IMAGE PROCESSING FOR MOVING OBJECTS

Wu Qing Xue

September, 1989

A thesis presented for the degree of
Doctor of Philosophy
in Electrical and Electronic Engineering
in the University of Canterbury
Christchurch, New Zealand

Abstract

This thesis reports research on aspects of medical imaging and general image processing.

As regards general image processing, techniques for estimating two-dimensional object motion (translation and rotation) from digital images are the main concern. Established techniques are reviewed, and new approaches are proposed. Several extensions are reported to the phase correlation (PC) technique introduced originally by Kuglin and Hines [1975] for estimating translation. The superior performance of PC as compared to that of conventional correlation, is shown to be related to the concept of phase dominance, a term quite recently introduced in this laboratory in the context of Fourier phase retrieval. Three modifications to PC are found to be necessary for it to be able to operate effectively on images exhibiting imperfections of the kind encountered in the real world. The ability of PC to estimate the translation of an object, when the latter also rotates, is examined and illustrated. A technique called *sequential predictive phase correlation* (SPPC) is introduced. It is particularly useful for estimating the translational motion of objects recorded in sequences of images.

Emphasis is placed on the importance of descriptors which permit object rotation to be estimated, at comparatively small computational expense, with the aid of angular correlation. Two descriptors, the bispectrum (equivalent to the triple autocorrelation) and the magnitude of the conventional spectrum, are examined in detail. The chief advantage of the bispectrum is that it is completely insensitive to translational motion of the object, even though it retains all of its phase information (thereby sharpening the peaks of angular correlations) in marked contrast to the spectral magnitude. The theoretical basis for motion estimation by bispectral analysis is presented. Numerical comparisons of rotation estimation by angular correlation employing the bispectrum and the conventional spectral magnitude, in situations where the level of noise and other contamination is appreciable, show that the performance of the former is significantly superior.

The main concern on the medical imaging side is with two aspects of digital coronary angiography. One of them relates to compensating for limitations of X-ray/video imaging systems. Screen brightness persistence, or lag, of video cameras typically employed in angiographic systems causes images acquired by them to be effectively multiple exposures, so that images of moving objects become significantly blurred. A method of compensating for the lag by weighted subtraction of successive images is presented. This method can achieve resolutions comparable to those

conventionally obtainable for images of static objects, but with an increase in the overall contamination level. Averaging with motion compensation, however, allows the original signal-to-noise ratio to be restored.

A second aspect of medical imaging examined and extended in this thesis is estimating the motion of coronary arteries for: 1) enhancing images for improving clinical diagnosis, and 2) assessing physiological function, especially coronary circulation. The key to this is shown to be appropriate compensation for arterial motion. New algorithms, based on the SPPC and lag removal techniques, are introduced. They are applied to the processing of clinical coronary angiogram sequences. The results presented herein show that these algorithms can estimate coronary arterial motion to useful accuracies. It appears that these algorithms could be usefully applied to the evaluation of coronary arterial narrowing due to coronary disease.

Aknowledgements

I am indebted to many people who have encouraged, supported, and advised me during the period of my Ph.D study.

Firstly I wish to thank my wife, Chao Ying, for her love and understanding, her consistent support of my study, as well as her contribution to the preparation of this thesis. I also thank her for our many discussions, some of which have proved to be extremely important for me.

My sincere thanks go to my supervisor, Professor Richard Bates, and co-supervisor, Dr. Phil Bones, who have directed my research enthusiastically and with enormous encouragement and tolerance. Not only have they provided excellent supervision, they have also made considerable effort to make my stay in New Zealand enjoyable.

I greatly value the support from many of my colleagues. I feel that without their help I could not have achieved as much as what I have. In particular, I thank Bruce McCallum, Peter Gardenier, and Bruce Davey for the enormous time they have spent in helping with various difficulties I have encountered since I started my Ph.D study. Several colleagues, Willian Thorpe, Andrew Elder, Bruce McCallum, Peter Gardenier, and Ross Murch, kindly assisted by proof-reading this document.

Two members of technical staff, Art Vernon and Lorance Hill (now retired) deserve special thanks for their assistance of various kinds.

I gratefully acknowledge the receipt of both a one-year scholarship from the Chinese Academy of Science and two-years of financial support from the New Zealand National Heart Foundation.

Contents

ERRATA

Page 1	line 9	There is good ...		
Page 4	line 28	...Computer Aided ...		
Page 6	line 35	... is a $J \times K$ ...		
Page 11	line 3	There is at least ...	line 25	... as a special ...
Page 18	(1.17)	... $\simeq -x\psi - b$		
Page 21	line 17	... is discussed.		
Page 22	(2.1)	lower limit $-\infty$	(2.2)	missing ')'
Page 22	(2.2)	lower limit $-\infty$		
Page 24	(2.12)	lower limit $-\infty$		
Page 26	(2.18)	lower limit $-\infty$		
Page 28	line 3	... combine the ...	line 18	... the Fourier space ...
Page 35	line 8	..., a resolution ...		
Page 38	line 14	... narrower than ...		
Page 45	line 4	... which varies from ...		
Page 50	line 1	... discussed.	line 8	Table 4.1 lists ...
Page 53	fig cap	line 4 ... (see (4.3)).		
Page 59	line 37	... a featureless ...		
Page 72	figure	**background**		
Page 74	fig cap	line 6 ... the AOI in ...		
Page 79	fig cap	line 6 ... to 4° rotation.		
Page 86	line 32	... α and $(u + \alpha)$		
Page 106	line 7	... of contrast material ...		
Page 115	line 11	... a quantum ...	line 39	To satisfy the sampling ...
Page 116	line 30	... (of the same ...	line 31	... from a digital ...
Page 123	line 6	... that can be encountered ...		
Page 139	line 4	... this is because the final ...		
Page 142	line 21	...disease and its ...	line 22	... coronary abnormalities.
Page 171	ref. 9	line 2 ... initial clinical ...		
Page 171	ref. 10	line 3 ... with conventional...		
Page 173	ref. 1	line 2 ... using finite ...		
Page 174	ref. 10	line 1 ... Video-densitometry ...		
Page 176	ref. 2	line 1 ... three dimensional ...		
Page 176	ref. 10	line 1 Blood flow determination ...		
Page 177	ref. 2	line 1 ... and Holcroff ...		
Page 177	ref. 7	line 2 (Review) Digital ...		
Page 180	ref. 3	line 1 Intravenous digital ...		
Page 180	ref. 4	line 2 ... of temporal ...		
Page 180	ref. 5	line 1 ... digital subtraction ...		
Page 180	ref. 8	line 1 ... three dimensional solids, ...		
Page 183	ref. 3	line 2 ... physiology with ...		

Preface

After completing my Sc.B in Physics in July 1982 in the Physics Department, Nanjing University of the People's Republic of China, I worked as a research assistant in Wuhan Institute of Physics, Academia Sinica in Wuhan, China, under the direction of Professor P. J. N. Wei, then the head of the Acoustical imaging Laboratory and Director of the Institute until 1985. With the help of Professor Wei, I obtained in late 1984 a one year scholarship from the Academy to study overseas. I was recommended by Professor Wei to Professor R.H.T. Bates with whom I started my Master of Engineering (M.E.) course in Electrical and Electronic Engineering in November 1985 at the Engineering School of the University of Canterbury, Christchurch, New Zealand.

During my first year in New Zealand I became involved in a cardiac imaging project at the Princess Margaret Hospital in Christchurch. This work has been jointly supervised by Dr. P.J. Bones, who was a Senior Fellow of the National Heart Foundation of New Zealand until February 1988 when he joined the University of Canterbury as a Senior Lecturer in Electrical and Electronic Engineering.

Towards the end of 1986, Dr. Bones successfully applied for a two-year research grant from the National Heart Foundation to study facets of cardiac angiographic imaging. This fortunately permitted me to participate more fully in this project than had at first seemed feasible. I was very pleased when Professor Bates suggested in December 1986 that I transfer from M.E. to Ph.D.

My Ph.D research work has been concerned with medical imaging and machine vision aspects of image processing. I have paid particular attention to methods of analysing image sequences with the goals of abstracting motion parameters and forming enhanced versions of objects whose appearances are much degraded in individual images.

During the past decade, image sequence analysis or dynamic scene analysis has emerged as a fast growing research area due to its potential application in fields such as robotics, biology, medicine, geoscience, etc. One of the major problems in image sequence analysis is registering particular objects in successive images of a sequence. Once solved, of course, this problem leads on naturally to the problem of estimating the motions of objects [Chellappa and Sawchuk, 1985, Overview of Part 3].

The analysis and processing of medical image sequences has been a very active field. This is especially true for X-ray image sequences obtained from video sources. Motion analysis and estimation are particularly important when a moving part of the

human body is being examined. All successful approaches to image sequence analysis rely heavily on pattern matching methods. In the human body, the heart is clearly the most mobile organ. In medical practice, relatively more attention is paid to the heart because of its vital significance to human health. *Angiography* is the technical term for a commonly invoked method for examining the circulatory system. It is well recognised by practitioners of angiography that current difficulties encountered when trying to enhance images of coronary arteries with the aid of image processing techniques are caused by the complicated motion of the heart.

Although many matching techniques based on processing in image space can be found in the literature, their Fourier-space counterparts have been comparatively neglected. This thesis attempts to redress the balance. It is further noted that, among the many well established image registration techniques, translational motion (or displacement) estimation has been extensively studied, but much less attention has been paid to rotational motion estimation.

Spectral analysis is emphasised in this thesis. New results are reported in four areas. The first relates to a technique for translational motion estimation called phase correlation by Kuglin and Hines [1975]. Several aspects of applying phase correlation to processing video image sequences are studied, leading to the establishment of a new form of phase correlation here termed *sequential predictive phase correlation* (SPPC), which is shown to be potentially useful in coronary angiography. The second area concerns the use of the Erlangen bispectrum (equivalent to what is often called the triple autocorrelation; cf. Bates 1987) as a rotation descriptor. This form of bispectrum has been christened *Erlangen* by Bates because it was developed primarily in the Physikalisches Institut of the University of Erlangen [Lohmann and Wirnitzer, 1984; Weigelt, 1977]. Possible applications of the bispectrum to machine vision tasks are suggested. The third area bears on video camera screen brightness persistence (often termed lag). A lag-compensation scheme, incorporating motion estimation by SPPC, is developed for processing coronary angiographic sequences. Some image sequence processing algorithms, based on SPPC and the above-mentioned technique for compensating for the camera lag, are presented. They are shown to be potentially useful for clinical diagnosis. The fourth area is related to reducing a particular type of electrical interference which degrades video signals of the kind generated by the angiographic imaging system in the Cardiology Department, Princess Margaret Hospital.

The first two chapters of this thesis mainly review motion estimation from the point of view of image processing and machine vision. Approaches to motion estimation are classified into those which are implemented in the *spatial* and *spectral* domains respectively. Chapter 6 reviews the fundamentals of X-ray angiographic imaging and techniques of digital image processing applied to angiography. Original material and new results are presented in Chapter 2 (see §§ 2.3.4 and 2.3.6), Chapter 3 (see §§ 3.2 and 3.3), Chapters 4 and 5, and Chapter 7 (see §§ 7.2 and 7.3).

In Chapter 1 the overall purpose of and the detailed problems associated with machine vision are introduced, emphasising image sequence analysis as an approach to achieving artificial vision. It is pointed out that, while in general three-dimensional analysis is necessary, two-dimensional analysis is sufficient in many practical appli-

cations. The general motion problem is examined mainly with a view to providing a background for subsequent discussions in later chapters. The rest of the chapter reviews spatial domain analysis of motion problems.

Chapter 2 concentrates on conventional (second order) spectral analysis for motion estimation. The Fourier transform and several of its properties which are essential to motion analysis are introduced. Spectral domain analyses of motion estimation are reviewed. The emphasis in this chapter is on phase-only approaches. The phase correlation technique for estimating translational motion is discussed in detail. The concept of phase dominance is invoked to explain the performance and characteristics of the phase correlation technique and its relation to the feature-based matching techniques discussed in Chapter 1. Two other methods, based on the conventional spectral magnitude and its cepstrum, are examined. Angular correlation techniques for estimating rotational motion, involving Fourier domain analysis, are established.

Higher order spectra, in particular the third order spectrum, or bispectrum, and its application to motion analysis are introduced in Chapter 3. Both its theoretical basis and its applications to electrical engineering and physical science are reviewed. The shift-rotation invariant properties of the bispectrum, on which its potential for rotation estimation is based, are explained. The bispectral angular correlation technique is also developed.

The implementations of both phase correlation and angular correlations are discussed, respectively, in Chapters 4 and 5. Practical considerations, algorithmic explanations and results for both correlation techniques are presented in these two chapters.

Chapter 4 discusses several practical aspects of processing video image sequences for object motion estimation and image enhancement. X-ray coronary angiographic image sequences are the main concern of this chapter. The effect of video camera lag on X-ray images of moving objects and its removal are examined in detail. The SPPC technique is established and its practical implementation is discussed. Several algorithms are introduced and illustrated with computational examples.

The practical implementation of angular correlation is discussed in Chapter 5. The performances, in the presence of appreciable levels of contamination, of rotation estimators based on the conventional spectral magnitude and the bispectrum are quantitatively compared. The advantages of descriptors (such as those based on the bispectrum) which incorporate phase information are emphasised. Various computational aspects of angular correlation techniques are explained.

Coronary angiographic imaging and image processing techniques are reviewed in Chapter 6. Cardiac function and disease, and the anatomy of the human heart, are discussed with emphasis on the coronary circulation. Diagnostic approaches to coronary disease, and the evolution of digital angiographic image processing techniques, are reviewed in detail. Commercial X-ray/video imaging systems are appraised. The mechanism by which conventional X-ray images are formed is discussed, as are the functions of major elements in the imaging chain, namely, image intensifier, video camera, and analogue-to-digital (A/D) converter. Resolution considerations and noise sensitivities for an entire imaging system are analysed. Various image pro-

cessing techniques commonly employed in digital angiography are classified under the categories of image enhancement and parametric (or functional) imaging. The problems associated with coronary arterial motion which have to be solved in order to successfully apply image processing techniques to imaging the coronary circulation are identified and discussed in detail.

A specialist hardware and software package, the Cardiology Angiographic Processing System (CAPS) developed by Dr. P.J. Bones and Dr. R.G. Lane (now with the Electrical Engineering Department of the University of Adelaide, Australia) at the Princess Margaret Hospital, is described in Chapter 7. A new approach to reducing a particular species of electrical interference, called spike noise, which often degrades angiographic images obtained from the X-ray angiographic system in the above-mentioned hospital is introduced. Motion estimation and image enhancement for clinical angiographic image sequences are discussed and illustrated.

Conclusions and suggestions for future research are given in chapter 8. During the course of this research, the following publications and presentations have been prepared:

Wu, Q.X., Bones, P.J. and Bates, R.H.T. (1989), 'Translational motion compensation for coronary angiogram sequences', To Appear on September's Issue of *IEEE Trans. Medical Imaging.*

Wu, Q.X. and Bones, P.J. (1987), 'Removal of spike interference from digitized angiographic images', *Australasian Physical & Engineering Sciences in Medicine*, Vol. 10, No. 2, Pp. 96–101.

Bones, P.J., Wu, Q.X. and Bates, R.H.T. (1987), 'Detection of translational motion in coronary angiogram sequences', In *Proceedings of the 10th Annual Conference of the Australasian College of Physical and Engineering Science in Medicine*, Auckland Medical School, August.

Bones, P.J., Wu, Q.X. and Bates, R.H.T. (1988a), 'On methods for detecting object translation in low-contrast image sequences', In *Proceedings of 3rd New Zealand Image Processing Workshop*, Massey University, Palmerstone North.

Bones, P.J., Wu, Q.X. and Lane, R.G. (1988b), 'Video-rate processing with conventional image acquisition hardware', In *Proceedings of the National Electronics Conference of New Zealand (NELCON '88)*, Christchurch, New Zealand.

Abbreviations

Selected abbreviations that appear in this thesis are:

CA	coronary angiographic (images)
CC	cross correlation
DFT	discrete Fourier transform
DSA	digital subtraction angiography
DSC	deterministic sign change
FFT	fast Fourier transform
LHS	left hand side
PC	phase correlation
RHS	right hand side
SAVD	sum of the absolute value of difference
SNR	signal-to-noise ratio
SPPC	sequential predictive phase correlation (maybe "super performance phase correlation")
SSC	stochastic sign change
SSDA	sequential similarity detection algorithm

Chapter 1

Inferring Object Motion From Images

This thesis is concerned with: 1) the analysis of object motion by means of image processing techniques; and 2) the application of motion estimation in imaging the coronary arteries of the heart. Since the earliest work on digital image processing most attention has been given to processing static images since this in itself demanded considerable resources of computing power. As the computational power has increased and as the field has expanded, more attention has been paid to the observation of objects in motion. This is particularly important since many imaging tasks involve observing objects in motion (it is clearly not possible to arrest the heart to take a picture!). There is good evidence also that the remarkable performance of the visual systems of many animals is, in part, due to the detection of motion [Cornsweet, 1971].

In this chapter, the concepts of machine vision and image sequence analysis are introduced in §1.1. §1.2 discusses motion analysis techniques in general. §1.3 presents the main mathematical techniques that have been developed for motion estimation.

1.1 Introduction

Human beings have sophisticated vision systems controlled by their brains. They perceive and interpret scenes and objects around them. It is clear that humans can quantitatively infer object motion to a certain extent. However, the human visual system is not capable of accomplishing unaided many important tasks, such as visualising internal human body structures as a physician does with the aid of X-ray [Hill, 1975] or ultrasonic [Wells, 1977] diagnostic equipment, or detecting distant targets in the manner of a radar system [Skolnik, 1962]. Moreover, nobody can accurately track moving objects without the assistance of measurement apparatus. For instance, it is impossible without technical aids to determine the speeds of vehicles on a road in order to monitor and control the traffic [Agin, 1979; Wolferts, 1974; Hartley, 1988].

Throughout history humans have long searched for ways of extending their visual abilities. Various "machine vision" systems, such as those mentioned in the previous paragraph, have been developed for different purposes. It is now common for a machine vision system to comprise an "eye" (the imaging part) and a computer (the "brain"). Recent research activity in machine vision has been mainly concerned with using computers to analyse and interpret perceived scenes or objects. This work is therefore often termed computer vision or computational vision [Barrow and Tenenbaum, 1981; Ballard and Brown, 1982; Rosenfeld, 1984].

Computer vision relies heavily on the use of image processing techniques, because what one perceives through an imaging system are only images representing the scenes or objects. As Rosenfeld and Kak [1982] explain, image processing tasks include: 1) digitisation and compression of images; 2) enhancement, restoration and reconstruction of images; 3) matching, description and recognition of images.

A modern imaging system, equipped with fast scanning electronics, is capable of capturing a sequence of pictures (called frames) at the video rate of 25 frames/sec or higher [Batchelor *et al.*, 1985]. Controlled by a computer, real time digitisation of sequences of 512×512 pixel images, and subsequently their storage in large frame buffer memories, is straightforward. The processing of image sequences plays an important role in various applications. Furthermore, image sequence analysis has been a fast growing field over the last decade [Chellappa and Sawchuk, 1985].

The major problems with which image sequence analysis and processing are concerned are enhancement [Huang and Hsu, 1981], restoration [Tsai and Huang, 1984], compression of the enormous amount of data present in sequences of images [Dubois *et al.*, 1981; Rocca, 1972], estimation of object motion, tracking, recognition and segmentation of objects. The wide spectrum of the applications includes target tracking, vehicle guidance, robot navigation and industrial automation, bandwidth compression of TV conferencing video signals, studying the motion of biological cells, processing of medical image sequences such as coronary angiograms (treated in this thesis), cloud tracking, traffic monitoring, satellite or airborne imagery for geographic mapping and earth resource analysis. Numerous references can be found from [Nagel, 1981; Nagel, 1983; Aggarwal, 1984; Aggarwal and Martin, 1983].

The treatment of the above mentioned problems in image sequence analysis, unlike ordinary image processing, is conditioned by the time varying nature of image sequences. Huang and Tsai [1981] conclude that one of the most important problems in image sequence analysis and dynamic scene analysis is displacement and motion estimation. In many image sequence processing problems, motion analysis is the key issue. For example, in efficient interframe coding using DPCM (differential pulse code modulation) in time [Gonzales and Wintz, 1977], motion estimation and compensation can potentially improve the efficiency significantly. In reducing noise in image sequences by means of temporal filtering, registration of the objects of interest from frame to frame is necessary (and note that registration is, in essence, equivalent to motion estimation). Finally, in tracking multiple targets (moving differently), motion estimation provides a powerful way of segmenting and identifying the individual targets. It is worth noting that Nagel [1983] comments that many frame-to-frame changes can be attributed to quite general phenomena such as the relative motion

between observer, scene components, and light sources.

Many applications involve only two-dimensional (2-D) analysis. For example, pictures of printed or written characters, line drawings, and sketches are inherently two-dimensional (although drawings can represent three-dimensional (3-D) information). Pictures of terrain taken from far away (e.g. from satellite altitudes) are essentially 2-D because the terrain relief is negligible at great distances [Bates and McDonnell, 1986; Kashef and Sawchuk, 1983; Svedlow *et al.*, 1978]. Conventional X-ray pictures (or shadowgrams) are projections of objects onto planes, i.e. they are 2-D "shadows" [Peters, 1973; Hohne and Bohm, 1983]. Photomicrographs taken at high magnification and shallow depth of field show only a thin slice of any microscope slide scene, since only this slice is in focus [Levine *et al.*, 1979].

On the other hand, pictures of scenes taken from relatively close by usually do require 3-D interpretation, in terms of perspective views of opaque solid objects. Object motion which involves rotation with respect to axes not perpendicular to the image plane needs to be estimated by 3-D analysis. Although a nonrotating rigid body, which only translates in the direction perpendicular to the image plane, can of course be interpreted by 3-D analysis, it may equivalently and often more conveniently be analysed in 2-D, provided appropriate scale changes are allowed for. Occlusion (i.e. the obscuration of one body by another, when the first moves behind the second) is a 3-D problem. However, in some situations, such as cloud tracking where the observation point is at a large distance, occlusion may be treated by 2-D analysis [Aggarwal and Duda, 1975; Aggarwal *et al.*, 1981].

This thesis is mainly concerned with problems of 2-D motion analysis, although the treatment of 3-D motion is briefly discussed in §1.2.1. Unless specifically noted otherwise, all discussion and analysis refer to 2-D problems.

1.2 Motion Estimation – spatial domain analysis

1.2.1 General Remarks

The types of object motion considered in this thesis are *rotation* and *translation*. It is assumed that *scaling* is an effect of 3-D motion for reasons which are made clear later. Furthermore, considering the mechanism of image formation, the object or objects of interest are usually recorded together with other unwanted components of scenes, which form non-uniform backgrounds. Often, the background is stationary during the observation of a scene. For example, buildings, fields and hills remain stationary when one monitors a moving vehicle. There are situations where non-stationary backgrounds exist. For example, trees may form non-stationary part of a background due to wind. A particular situation involving non-stationary backgrounds considered in this thesis is X-ray imaging of the human body, where different parts may move differently. Therefore, the part of interest is overlaid on other moving parts which form an undesirable non-stationary background. The transparency of the human body to X-rays makes the background very difficult to deal with. On the other hand, images of patterns obtained in the laboratory under ideal conditions, or

computer-generated images of objects, may have uniform backgrounds.

Based on the above considerations, most motion problems occurring in the various applications that are currently of technical interest, can be grouped into three types:

1. *Single object motion* includes: (a) single rigid object motion; (b) multiple rigid objects moving without occlusion so that each may be treated separately; (c) camera motion (this is a type of single rigid object motion because the whole scene translates or rotates). In situations a) and b), backgrounds tend to be comparatively flat and featureless (i.e. close to uniform).

2. *Multiple-object motion* includes: (a) multiple rigid object motion involving occlusion; (b) deformable moving objects, for example X-ray images of the human heart; (c) objects moving against distinctly non-uniform backgrounds.

3. *3-D motion* includes: (a) a single rigid body moving in 3-D with arbitrary rotation and translation; (b) multiple rigid bodies moving in 3-D, involving occlusion.

While group 1 can usually be easily dealt with using rigorous generally applicable approaches which are described later, the treatment of group 2 tends to be application oriented, i.e. subject to significantly different practical constraints in each application. However, the analysis of single object motion forms the basis for the treatment of multiple-object motion. In other words, the former can be seen as the idealized situation (although it relates to numerous practical applications, for example, terrain mapping in remote sensing [Kashef, 1985]). The latter can be approached by approximation, or by detecting features of objects and using special transformations (such as those described in §1.2.3) so that each object can be treated separately.

Successful analysis of 3-D motion is based on having an *a priori* model [Roberts, 1977] of the expected object, or on *a priori* knowledge of corresponding features in an image sequence [Huang *et al.*, 1986; Liu and Huang, 1988; Huang and Tsai, 1981].

Object or scene modelling is particularly important but is also difficult [Binford, 1982; Rosenfeld, 1984]. One approach is to use Computer Aided Design (CAD) based vision systems [Bhanu, 1987; Bhanu and Ho, 1987]. In cases where no useful models can be devised, various techniques have been developed for interpreting 3-D object or scene information [Ballard and Brown, 1982]. Some of these techniques involve invoking image features such as shading, occlusion, size, shape and perspective to infer depth information [Shirai, 1978; Horn, 1977]. Others invoke range finding devices to map the object in depth [Sampson, 1987; Waag and Gramiak, 1976]. Stereo analysis is useful for determining 3-D positions [Shirai, 1978; Kim and Aggarwal, 1987]. A recent proposal by Murch *et al.* [1988] (cf. Heel 1986) involves using silhouettes of an object to infer its 3-D shape and orientation with the help of the projection theorem, as employed in computer-assisted tomographic imaging [Bates *et al.*, 1983; Garden, 1984]. 2-D methods, such as matching and differentials (described in §§ 1.2.2, 1.2.3 and 1.2.4), can be employed to locate corresponding features in image sequences for subsequent 3-D motion analysis [Huang and Tsai, 1981]. *Optical flow*, which represents time-varying imagery by assigning to each pixel a velocity

vector describing its temporal change in the image plane (or for human vision, in the retina field) [Horn and Schunck, 1981; Ullman, 1979; Neumann, 1984], provides powerful clues to the location and orientations of objects relative to an observer.

It is important to recognise that images are often recorded with various unavoidable distortions and subject to certain practical constraints. One kind of distortion is the random fluctuation endemic to image formation mechanisms, such as the additive thermal electrical noise from electronic devices (e.g. TV cameras) [Macovski, 1983; Shaw and Bassano, 1982; Kruger *et al.*, 1981] and the inherent quantum fluctuation of radiographic images [Macovski, 1983].

Another group of distortions is caused by systematic degradations due to image formation geometry as well as the optical limitations of the imaging system. Examples are perspective, pincushion and barrel distortions. These types of distortions are known collectively as geometrical distortion [Rosenfeld and Kak, 1982]. Rectification of various geometric distortions is considered by Bates and McDonnell [1986 §47; cf. Rosenfeld and Kak 1982; Kashef 1985].

Other forms of distortions also exist. For example, variable lighting conditions may form a kind of distortion, unless a radiometric model can be constructed. Or again, the finite size of any real-world image (caused by the unavoidable finiteness of the field of view) may exclude part of the object of interest, if it is in motion (this is particularly apposite for image sequences). A particular distortion considered in Chapter 4 of this thesis is the blur caused by the photoconductive persistence of TV cameras. This effect is conventionally termed *camera lag* [Davison and *et. al*, 1977].

Ideally, before motion estimation is attempted, the above distortions should be corrected. However, it is often impossible to remove some of them in practice. On the other hand, most methods for estimating object motion have certain distortion suppression properties (see §§1.2.2 and 1.2.3). Various distortion-tolerant methods have been employed in matching geometrically distorted scenes and may provide useful information for motion analysis. For example, the ideas of "rubber sheets" [Widrow, 1973] and "spring connected" subtemplate matching [Fischler and Elschlager, 1973] can give at least the approximate location of a scene or object which one wishes to track. Relaxation techniques [Davis and Rosenfeld, 1976; Rosenfeld and Kak, 1982] are also employed in subtemplate matching to eliminate bad matches by iterative schemes. Note that a subtemplate is a part of a template small enough that it can be matched well under distortion.

In this thesis, two general approaches to motion analysis are taken: *spatial domain analysis* and *spectral domain analysis*. Most attention is paid to the spectral approach, which is discussed in Chapters 2 and 3. The spatial approach is examined in the remainder of this chapter.

There are basically three types of spatial domain analysis:

1. intensity based matching [Aggarwal *et al.*, 1981; Rosenfeld and Kak, 1982];
2. feature based matching [Aggarwal *et al.*, 1981; Rosenfeld and Kak, 1982];
3. the differentials method [Cafforio and Rocca, 1983].

Matching techniques (i.e. types 1 and 2) have been heavily used in pattern recognition. Typically, a template, which is the computer representation of an object, and which can comprise either intensity variations over an image of the object or features of the object itself, is matched to the object which one expects to exist within the images which are being examined [Aggarwal *et al.*, 1981; Rosenfeld and Kak, 1982]. The first two types of motion analysis are discussed, respectively, in §1.2.2 and §1.2.3. The differentials method [Cafforio and Rocca, 1983], which estimates object motion on the basis of temporal intensity variations within images, is discussed in §1.2.4.

1.2.2 Intensity Based Matching

In this method a template is constructed from the intensity variation occurring over a small but strongly featured area in one (reference) image of a sequence [Aggarwal *et al.*, 1981]. If the object or feature is sufficiently small, the template may enclose it entirely in the reference image. The template is matched with other frames in the sequence. Correct matching accurately gives the positions or the relative translation and rotation of the feature or the object in the image sequence. A measure of actual object motion can be obtained by calibration between the actual scene and the images. For example, in order to determine motion of a cell from a microscope-video system [Levine *et al.*, 1979], the calibration between the scale under microscopic magnification and the digitised image samples is needed. However, in applications such as TV image coding [Robbins and Netravali, 1983], temporal image enhancement [Huang and Hsu, 1981], etc., only the shifts or rotations in images from frame-to-frame in the sequence need to be determined. Matching usually requires that the moving object remains invariant in its attributes such as size, shape and grey levels [Nagel, 1981].

In general, the goodness of match, sometimes referred to as the similarity measure [Barnea and Silverman, 1972], can be represented by a cost function. Let two images of an object obtained at different times be denoted by f and g respectively (where f is the template). Then, the cost function is defined to be

$$Cost = \text{Match}(f, g) \tag{1.1}$$

The *cost* is seen to be a function of the spatial coordinates and it usually has its maximum value or minimum value at the position of the best match obtained. In the matching procedure, every possible position in the image to be matched is usually considered, i.e. the entire space of the cost function is examined. This is often computationally expensive. If rotation is considered, *cost* is a function of angle as well as position. The large required computational effort generally limits its use to translation. The search scheme is illustrated in Fig. 1.1. Suppose that f is a $J \times K$ pixel template and g is an $M \times N$ pixel area to be matched to f. It is assumed here that $M > J$ and $N > K$. In general the *cost* is computed for all $(M - J + 1)(N - K + 1)$ possible translations of f, with respect to g, needed to find the displacement that gives the best match.

The most widely used matching techniques are based on the cross-correlation (CC) and Euclidean distance measures [Aggarwal *et al.*, 1981; Rosenfeld and Kak, 1982].

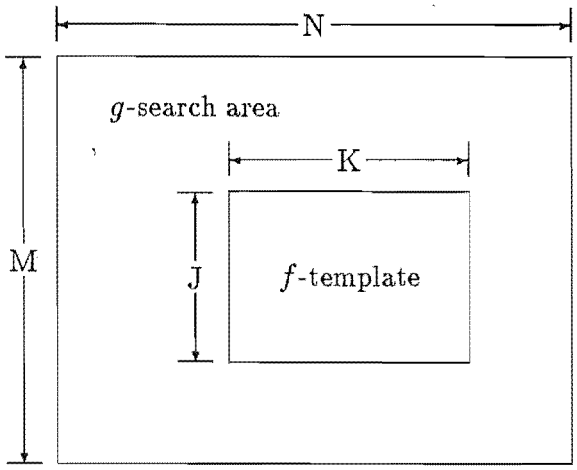


Figure 1.1. Illustration of the matching scheme and the relationship between the search area g and the template f.

In practical implementations of Euclidean distances, only the sum of the absolute value of difference (SAVD) is popular because it is relatively inexpensive computationally. Recently, Venot *et al* [1984; cf. Venot and Leclerc 1984] have suggested a new similarity measure which is here called the sign change measure. Two variants (i.e. the stochastic sign change (SSC) criterion [Venot *et al.*, 1984] and deterministic sign change (DSC) criterion [Venot *et al.*, 1984]), have been constructed. For both of these the *cost* is the number of sign changes in the difference image between f and g. A match is deemed to have been achieved when the *cost* is a maximum (see §1.3.2).

The use of CC algorithms is usually justified by their ability to maximize the signal-to-noise ratio (SNR) in the presence of (many kinds of) noise at the position of best match [Pratt, 1974; Rosenfeld and Kak, 1982]. They can also be justified on the grounds of minimum-error decision making [Rosenfeld and Kak, 1982]. Furthermore, they are implementable by analogue-optical means [Goodman, 1968; Casasent, 1981].

One disadvantage of CC algorithms is that the computation requirement is very demanding because of the large number of multiplications needed. When the FFT algorithm [Cooley *et al.*, 1967] is employed for computing the cross-correlation, the time can usually be considerably reduced, but at the cost of larger memory requirements [Barnea and Silverman, 1972].

Another difficulty with CC algorithms is that the cross-correlation function may be very broad, making detection of the peak difficult. This is particularly serious when the images being examined are devoid of sharp features. Pratt [1974] has described the use of a *statistical correlation measure* for avoiding this difficulty. His method is effectively of the feature-based kind because he, first, preprocesses the

images with a filter (constructed from the statistical correlation measure) that effectively performs edge extraction and, then, compares the edge outlines.

The size of the template also has an affect on the outcome of a CC algorithm. If it is too small, CC tends to lose its advantage of suppressing statistical noise. If too large, it is sensitive to systematic error when non-statistical degradations, such as geometric distortions, exist [Smith *et al.*, 1979; Leo and Gerson, 1979].

The SAVD algorithm registers f with g by minimum mismatch, i.e. at the position of best match, the sum of absolute values of differences for all points in f and the corresponding part in g reaches a minimum. It is mathematically much simpler and computationally much cheaper than the CC algorithm. However, one disadvantage is obvious: even in the absence of any non-statistical distortion, subtraction leaves only the reinforced random noise from the template and the area being searched. This is likely to make the determination of minimum mismatch erroneous.

The sign change measures [Venot *et al.*, 1984] are reported to have sharper registration peaks than the CC and SAVD. The stochastic sign change (SSC) [Venot *et al.*, 1984] criterion is developed especially for matching images with significant random fluctuations, such as nuclear medicine images. Conversely, the deterministic sign change (DSC) criterion [Venot *et al.*, 1984] is suggested for matching low-noise images, such as high quality x-ray images. Details of algorithms for sign change measures are given in §1.3.

In order to reduce the large amount of time needed for conventional matching algorithms to compute all mismatches before determining the displacement for a match, Barnea and Silverman [1972] have developed the sequential similarity detection algorithm (SSDA), which is based on the SAVD algorithm. The basic idea is that the accumulated sum of differences for a particular shift is compared with a fixed threshold before all points in f are examined, so that poor candidate shifts can be discarded quickly. The final displacement is chosen from the shift corresponding to the largest number of pixels examined for the given threshold. For the example presented by Barnea and Silverman [1972], SSDA is about 50 times faster than a CC implementation using the FFT algorithm.

Another approach to increasing the speed of computation has been the development of coarse-to-fine algorithms [VanderBrug and Rosenfeld, 1977; Rosenfeld and VanderBrug, 1977]. One type of these computes *cost* at a coarsely spaced grid of pixels and then searches for better matches in the vicinity of those grid pixels where the *cost* indicates high matches between f and g [Nagel and Rosenfeld, 1972]. Another type employs a subtemplate [VanderBrug and Rosenfeld, 1977; Aggarwal *et al.*, 1981] for coarse search. Still another uses a template and images which are reduced-resolution versions (e.g. by detecting every second pixel, horizontally and vertically) of the originals [Rosenfeld and VanderBrug, 1977].

It is advantageous to combine several of the methods discussed above. For example, combining CC and SAVD through the use of fast matching algorithms, as suggested by Kashef [1985], can be expected not only to give the same accuracy as the CC algorithm but also to reduce computational effort dramatically. The CC algorithm should be used only for the fine search, i.e. the *cost* of cross-correlation should

computed in the vicinity of the position where the *cost* of SAVD is a minimum.

When tracking an object in an image sequence, computational cost may still be large even with the use of the fast techniques discussed above. However, one can use the object's position in one frame to predict its approximate position in the next frame. Therefore, once the position has been found in the first frame by exhaustive search, one may only need to examine a portion of each subsequent frame.

Variation in illumination conditions during the recording of an image sequence causes inconsistencies in grey-level distributions of the images in the sequence. When no *a priori* information is available concerning reflectivities of objects and the illumination over them, direct intensity based matching techniques are sensitive to changes in illumination conditions [Huang and Tsai, 1981; Svedlow *et al.*, 1978]. The intensity based matching techniques are also sensitive to occlusion because the portion of an image being searched for may be occluded due to motion (this can be avoided by choosing a small template, if only partial occlusion occurs during the recording of an image sequence). The feature-based matching described in the next subsection is inherently insensitive to illumination conditions and often functions satisfactorily in the presence of occlusion.

1.2.3 Feature Based Matching

Rather than represent an image of an object purely as a 2-D function of intensity, it is frequently possible and desirable to describe those features of the object which distinguish it from other objects [Nagel, 1983]. Examples of local features are edges and curves (straight or curvilinear) and their lengths, corners or vertices, etc [Aggarwal and Martin, 1983]. When the scene contains only simple binary patterns, global features such as centroids, areas, major and minor axes (for moment of inertia) may be considered [Aggarwal and Martin, 1983]. The features and their positional relationships are termed *invariant* or *quasi-invariant* entities in a depicted scene by Nagel [1983]. They are relatively insensitive to overall intensity distribution changes caused by, for example, variations in illumination conditions. It is therefore attractive to use the features for matching, and thus for determining the object motion [Huang and Tsai, 1981].

Another advantage of feature-based matching is its ability to deal with occlusion in some situations. Firstly, in some situations, certain features of an object are not occluded during the recording of the image sequence, so that the motion of the object can be determined from them. Secondly, the relative motion of two features in an image sequence can indicate the existence of two objects which occlude each other (because of occlusion, two objects may merge together during an image sequence). This idea can be used to segment an object from its surroundings [Martin and Aggarwal, 1979].

Feature-based matching involves processing in two steps. In the first step, features of objects are abstracted from scenes. It can be adequate to identify primitive features such as positions of all pixels in an edge segment. It is of course better to identify other descriptive values besides location. For example, a segment of an edge or

curve is characterised by orientation and curvature. Another example is a corner of a polygon where the feature comprises the angle, sides and the location of the vertex [Aggarwal and Duda, 1975]. Various segmentation techniques [Rosenfeld and Kak, 1982, Ch.10] are useful for feature detection. Low-level operations such as edge or curve detection are popular with derivative operators often being used. Usually, pixels detected as parts and non-parts of a feature are labelled 1 and 0 respectively. Here, a non-zero pixel is termed a feature point or simply a point. The second step is to match features or feature sets in different frames in order to determine the object motion. Rosenfeld and Kak [1982] describe a general approach to this kind of feature matching. Each feature in the first set is paired off with every feature in the second set, so as to determine the displacements under which they coincide. If there are m features in each set, this gives m^2 displacements; m of these should be the same, corresponding to the correct displacement, while the others should be quasi-randomly distributed, so that one should have a detectable peak to match in the displacement space. This is equivalent to cross-correlating two binary images, with pixels labelled 1 at the positions of the features. The computational cost of this process is proportional to the square of the number of feature points. If they are sparse, the cost may be much less than that of matching image intensities. This technique can also be used to match objects with rotational motion. Suppose that two sets of edge points are given with associated orientations. One can determine the orientation difference of each pair of edge points; this gives m^2 differences, m of which should be the same, yielding a match peak. This corresponds to one-dimensional cross-correlation of the orientation histograms corresponding to the two sets of edge points.

There are many examples in the literature of the approach to feature-based matching described in the previous paragraph. Ballard [1981; cf. Aggarwal *et al* 1981] describes a system which uses the generalised Hough transform [Hough, 1962; Duda and Hart, 1972; Sklansky, 1978] for mapping feature points into an accumulator space, in which each value is the number of feature points in image space characterized by a given set of parameters. The peak in the accumulator space can indicate the position and orientation of the object characterized by the set of feature points. For example, let $B = \{(x_i, y_i)\}_{i=0}^{n}$ be the boundary points of an object and $P = (x, y)$ be any point in the image space, then the generalised Hough-representation of the boundary B using P, $H(B, P)$, is a sequence of vectors $\{(x - x_i, y - y_i)\}_{i=0}^{n}$ (the parameter set). When a set of feature points in one image is so transformed, the same set can be searched for (corresponding to the same set of parameters) in a second image which contains the same object. Through the search, the object translation and rotation are determined. This system has been applied to cloud tracking, and it can be generalised to multiple object motion and can deal partially with occlusion. Martin and Aggarwal [1979] have studied matching scenes containing curvilinear figures. They employ chain code representations of edge points and map the edge points into an *angle vs arc length* graph, i.e. the $\psi - s$ function. The gradient of a particular graph segment is equal to the curvature of the corresponding edge segment. The two graphs from a pair of images are compared and the positions and orientations of objects can thus be estimated. This system is particularly designed for dealing with occlusion (i.e. two or more objects merge to appear as a single object) and has been applied to scene segmentation through motion. The idea is that by assuming object rigidity,

a group of joined edge segments, which preserve their curvatures and do not break apart in consecutive frames, can be treated as features of one object.

There is at least one example in which the feature based matching technique is invoked in combination with the intensity based matching technique. Goshtasby [1985] has described template matching employing moments. In this method, the template lies in a circular window which is matched to the search area (see Fig. 1.1). Because of the window being circular, the central rotation-invariant moments [Goshtasby, 1985] can be determined (it is not true when the window is rectangular since a rotation can change the content of a given window). Subsequently, the position of the content of the template inside the search area can be determined by matching several central rotation-invariant moments, even when the content of the template is rotated with respect to that of the search area.

In general, the computational cost of feature matching is higher then direct intensity matching because of the need for scene segmentation and description [Aggarwal *et al.*, 1981].

1.2.4 Differentials Method

The third method of motion estimation by spatial domain analysis involves detecting moving objects directly on a pointwise basis, i.e. discretely spaced mathematical points are utilized for estimation. This is often termed the *method of differentials* [Huang and Tsai, 1981; Cafforio and Rocca, 1979; Cafforio and Rocca, 1983; Limb and Murphy, 1975]. The key idea is to relate the difference between two successive frames due to object motion to the spatial intensity gradient of one of the two frames. First order Taylor's series expansions are usually invoked, as approximate mathematical relations between the frame-to-frame changes and the spatial gradient (see §1.3). Most authors have only considered translation or displacement estimation in 2-D problems. Huang and Tsai [1981] treat 2-D rotation as a special case of 3-D motion. A general 2-D approach including rotation is formulated in §1.3.

The expansion of a function by first-order Taylor's series is exact only when the function is linear. Therefore, the differentials method for estimating motion is based on linear approximation. The use of linear approximation requires that, within the period of recording two consecutive frames, the object of interest moves (by translation and rotation) over regions much smaller than the size of object. This is necessary for the approximation to be useful because the object intensity variation within a small region may be considered to be approximately linear. Furthermore, in order to compute the intensity slope faithfully (as is needed for Talor's expansion), it is best to choose pixels in regions having smooth shading (intensity varies approximately linearly) if possible. Consequently, one advantage of this method is that it involves much less computation than ordinary matching, because usually only very small portions (small numbers of pixels) of the images are considered. For images exhibiting multiple-object motion (so that deformation and occlusion may occur), one need examine only those regions which appear to be translated and rotated. This avoids the complexity of multiple-object motion, i.e. it bypasses deformation or occlusion. It represents another advantage of the differentials method.

It is thus seen that an important aspect of the differentials method is that it allows convenient choices for the positions of pixels for estimating motion. So, it relies on the experience of the human operator.

It is desirable to use more than the minimum allowable number of pixels to find the solution in the least-squares sense [Huang and Tsai, 1981] in order to reduce the influence of noise. One may solve simultaneous equations arising from a number of pixels in a common neighbourhood through the use of the pseudoinverse technique [Albert, 1972; Ballard and Brown, 1982] which is, in essence, equivalent to least-square regression. Huang and Hsu [1981] have studied the use of pseudoinverse methods for the differentials method.

The differentials method can be directly extended to solve 3-D motion problems for which feature correspondences can be established for two or more frames [Tsai and Huang, 1981; Tsai *et al.*, 1982; Feng and Huang, 1984; Liu and Huang, 1988]. This is done by considering the 3-D geometry of the object. Therefore, variations in the total intensity of the 2-D image, caused by 3-D motion, must be accounted for. Point correspondences (critical pixels such as corners, sharp peaks in images) [Huang and Tsai, 1981; Huang *et al.*, 1986] or straight line correspondences (straight edges or other lines) [Yen and Huang, 1983; Liu and Huang, 1988] have been considered. 3-D changes may also be related to associated 2-D changes directly [Huang and Tsai, 1981].

It is worth mentioning that the computation of *optical flow*, the velocity vector field, relies on the differentials method described above [Horn and Schunck, 1981; Adiv, 1985; Ballard and Brown, 1982; Neumann, 1984]. The method of differentials is formulated in §1.3.2

1.3 Mathematical Algorithms

This section discusses some of the common mathematical formulations encountered in spatial motion analysis. Scene segmentation and description techniques are not considered here, and readers are referred to [Rosenfeld and Kak, 1982; Ballard and Brown, 1982] for a discussion of these topics.

Coordinate transformations between two Cartesian systems are recalled in §1.3.1 for later convenience. §1.3.2 presents some of the similarity measure algorithms invoked in matching procedures and describes the algorithm of the method of differentials.

1.3.1 Coordinate Transformations

This subsection first introduces particular coordinate systems and then discusses two basic coordinate transformations: translation and rotation [Rosenfeld and Kak, 1982]. These transformations are of primary importance for 2-D object motion. Attention is mainly paid to the 2-D situation here for the reasons given in the last paragraph of §1.1.

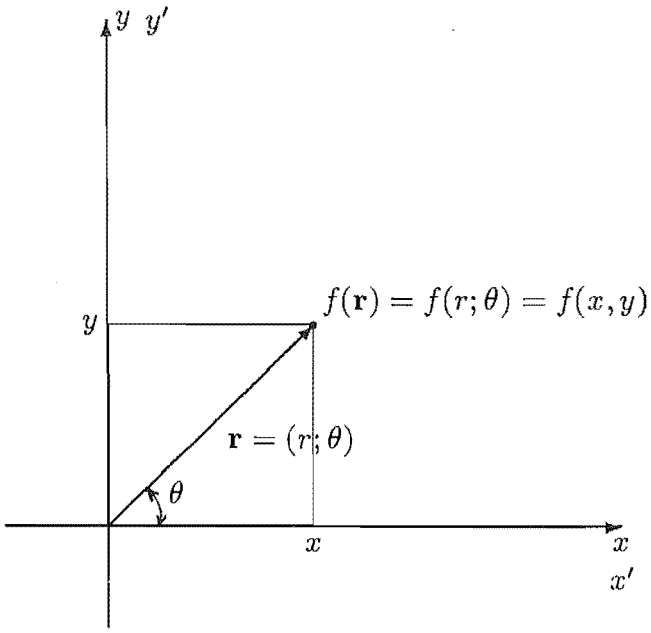


Figure 1.2. Cartesian and Polar coordinates systems.

The Cartesian coordinates x, y and the polar coordinates $r; \theta$ are illustrated in Fig. 1.2. It is assumed that the x, y-plane (or, equivalently, the $r; \theta$-plane) coincides with the image plane. The z coordinate is only mentioned in the analysis when 3-D aspects have to be explicitly taken into account. The coordinates (x, y) and $(r; \theta)$ are also represented by the position vector $\mathbf{r}$, which is hereafter called the spatial coordinate vector. Any 2-D function, such as $f = f(x, y) = f(r; \theta) = f(\mathbf{r})$, of the image-plan coordinates, whether positive or even complex, is called an image. However, most images relating to the practical situations discussed in this thesis are real and positive.

Let (x, y) and (x', y') be two Cartesian coordinate systems in the image plane. It is assumed that (x', y') is attached to a rigid object of interest and (x, y) is fixed in the image plane. Therefore, object motion can be envisaged as the (x', y') system moving with respect to the (x, y) system. It is also assumed that the origins of the two systems may be chosen to coincide (see Fig. 1.2) at a certain time. The task here is to find the relation between the two systems at any time when the object translates and rotates, that is: given the coordinates (x', y') of a point P for the moving system, the coordinates (x, y) of P for the fixed system are to be found.

Fig. 1.3 illustrates a *translation* by (a, b). The required transformation is given by

$$x = x' + a, \quad y = y' + b \tag{1.2}$$

A *rotation* of an object (counterclockwise about the origin of the (x, y) system)

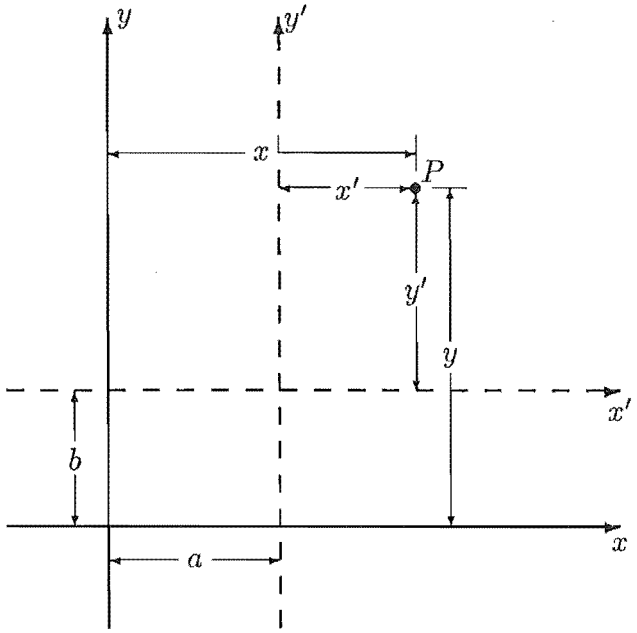


Figure 1.3. Coordinate transformation: a translation by (a, b)

through an angle ψ, which is shown in Fig. 1.4, is characterised by

$$\begin{aligned} x &= x' \cos\psi - y' \sin\psi \\ y &= x' \sin\psi + y' \cos\psi \end{aligned} \tag{1.3}$$

The rotation can also be represented by the dyadic $\mathbf{R}$, which is defined by [Morse and Feshback, 1953, §1.6]

$$\mathbf{R} = \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix} \tag{1.4}$$

The translation and rotation operations do not commute, for any given rotation centre, which is defined as that point where the axis of rotation intersects the image plane perpendicularly. In other words, the result of the two operations depends on the order in which the operations are performed (see Fig. 2.3). This can be readily confirmed by alternatively performing each of the two operations first. The origin of the (x, y) system is chosen for convenience as the rotation centre. Firstly, let the object be rotated through an angle ψ and then be translated by (a, b). The new coordinates are

$$\begin{aligned} x &= x' \cos\psi - y' \sin\psi + a \\ y &= x' \sin\psi + y' \cos\psi + b \end{aligned} \tag{1.5}$$

When these operations are performed in the opposite order, for the same trans-

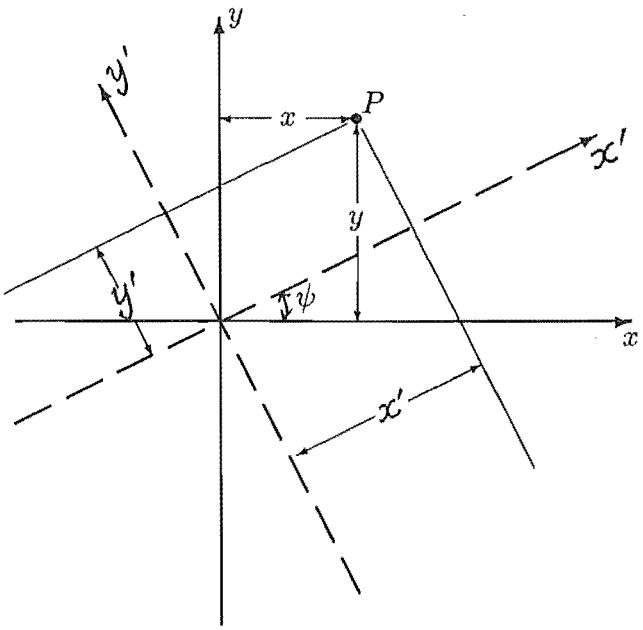


Figure 1.4. Coordinate transformation: a rotation through an angle ψ.

lation and rotation, the equations become

$$\begin{aligned} x &= (x' + a)\cos\psi - (y' + b)\sin\psi \\ y &= (x' + a)\sin\psi + (y' + b)\cos\psi \end{aligned} \tag{1.6}$$

Both expressions (1.5) and (1.6) are needed later. Note that (1.6) can be interpreted in the same way as (1.5), that is with translation following rotation, but with a different translation, the components a' and b' of which are here called equivalent displacements (see Fig. 2.3(a)). The latter are given in terms of a and b by

$$\begin{aligned} a' &= a\cos\psi - b\sin\psi \\ b' &= a\sin\psi + b\cos\psi \end{aligned} \tag{1.7}$$

1.3.2 Similarity Measures and Differentials

Motion estimation techniques are described in some detail in §1.2 without recourse to mathematical argument. This section presents the mathematical formulations which underly most of the techniques discussed in §1.2 [Huang and Tsai, 1981; Rosenfeld and Kak, 1982; Aggarwal *et al.*, 1981; Kashef and Sawchuk, 1983; Pratt, 1974]. It is assumed in all cases that it is valid to treat the discrete data set representing each image as samples of a continuous 2-D function on a regularly spaced grid. It is furthermore assumed that the samples are sufficiently closely spaced to avoid any significant aliasing of the information contained in images. Therefore, continuous

functions are mainly invoked in what follows. Discrete (or sampled, or pixellated) functions are only mentioned when the argument absolutely demands it.

Minimum Distance Measures

Various distance measures are defined by

$$D(\xi,\eta) = \iint_{\mathcal{R}} |f(x,y) - g(x+\xi, y+\eta)|^n dx\, dy \tag{1.8}$$

where $\mathcal{R}$, the region of integration, has the same size as the template f and $n = 1, 2, 3.....$ A particular case is the SAVD or the Euclidean distance for which $n = 1$. The coordinates (ξ, η) represent the displacement space, which is searched for a point (a, b) such that $D(a, b)$ is a minimum. Ideally, (a, b) indicates the correct measure of translation.

Cross Correlation

When (1.8) is expanded with $n = 2$ (assuming f and g are both real and positive), it gives

$$\begin{aligned} D(\xi,\eta) \;=\; & \iint_{\mathcal{R}} f^2(x,y)\, dx\, dy \\ & + \iint_{\mathcal{R}} g^2(x+\xi, y+\eta)\, dx\, dy \\ & - \iint_{\mathcal{R}} f(x,y)\, g(x+\xi, y+\eta)\, dx\, dy \end{aligned} \tag{1.9}$$

The first term on the right hand side (RHS) of (1.9), i.e. $\iint f^2(x,y)dxdy$, is a constant because f is the given template. If the second term on RHS of (1.9), i.e. $\iint g^2(x+\xi, y+\eta)dxdy$, is also a constant, the minimum distance for $n = 2$ corresponds to the last term on RHS of (1.9), i.e.

$$C_{fg}(\xi,\eta) = \iint_{\mathcal{R}} f(x,y)\, g(x+\xi, y+\eta)\, dx\, dy \tag{1.10}$$

being a maximum. (1.10) is just the conventional cross-correlation criterion for template matching.

However, the conventional CC process does not guarantee that its maximum is the expected match because the second term on the RHS of (1.9), i.e. $\iint g^2(x+\xi, y+\eta)\, dx\, dy$ usually varies with (ξ, η). Regions of high intensity in g may bias the expected maximum match (which corresponds the minimum distance for $n = 2$), especially when the template f is small. One way of overcoming this difficulty is to use the normalized cross correlation, as discussed by Kashef [1985; cf. Rosenfeld and Kak 1982]. Because $\iint f^2(x,y)\, dx\, dy$ is always a constant, the form of the normalized cross correlation is

$$\tilde{C}_{fg} = \frac{C_{fg}}{\left[\iint_{\mathcal{R}} g^2(x+\xi, y+\eta)\, dx\, dy\right]^{1/2}} \tag{1.11}$$

It is worth noting that the cross correlation criterion can be obtained by making use of the Cauchy-Schwarz inequality [Rosenfeld and Kak, 1982], which is effectively equivalent to the Euclidean distance measure.

When invoking the correlation theorem [Bates and McDonnell, 1986] for computing the cross correlation through the use of the FFT algorithm, the sizes of the template f and the image g each have to be enlarged by packing them with zeros, so that their sides become the sums of the sides of the template and the image (so that aliasing in the correlation domain can be avoided) [Gonzales and Wintz, 1977]. Therefore, if f contains $J \times K$ pixels and g contains $M \times N$ pixels, the enlarged images should contain $(J + N) \times (K + M)$ pixels. This significantly increases the computer memory requirement.

There is another aspect of the use of FFT algorithms that needs to be considered. In order to invoke the FFT efficiently, the number of pixels in both the image and its spectrum is required to be either a power of 2 or prime-factorisable [Cochran *et al.*, 1967; Cooley *et al.*, 1967; Burrus and Parks, 1984]. This, in turn, further tends to increase the memory requirement.

The Stochastic Sign Change (SSC) and Deterministic Sign Change (DSC) Criteria

As already discussed in §1.2.2, the SSC and DSC criteria for matching the template f with the image g are the maximum numbers of sign changes in the difference image of f and g. The justifications for these criteria are provided by Venot *et al.* [1984] and Gibbons [1971].

In order to invoke the SSC criterion, it is assumed that the template f and the part of g, having identical features to f, differ only because of noise which is considered to be additive, of zero-mean and having a symmetric probability density function. Let d denote this difference image. Therefore

$$d(i,j) = f(i,j) - g(i+h, j+l) \tag{1.12}$$

When f and g match perfectly, say at $h = a$ and $l = b$ of the displacement space, the values of d should exhibit only random fluctuations, statistically similar to those described previously, except having a variance which is the sum of the variances of f and g. Therefore, the values of d at $h = a$ and $l = b$ can be either positive or negative. There is also an equal chance (0.5 probability because of symmetry) for each sign being either $+$ or $-$.

The Stochastic Similarity Measure based on the SSC criterion is defined as,

$$S(h,l) = \mathrm{SSC}[f(i,j) - g(i+h, j+l)]_{i,j}^{J,K} = \mathrm{SSC}[d(i,j)]_{i,j}^{J,K} \tag{1.13}$$

where J, K are the side lengths of the template f, and $SSC[d(i,j)]_{i,j}^{J,K}$ is the total number of sign changes in the difference image d calculated by scanning the image line-by-line or column-by-column. For example, in the sequence $+, -, -, -, +, -$, there are 3 sign changes. As in other similarity measures, the displacement space

(h, l) is searched for the point (a, b) by maximising $S(a, b)$. Ideally, (a, b) indicates the match position, i.e. the translation.

It should be emphasized that the SSC criterion is based on the assumption that a significant random fluctuation level exists in the images. This approach has been applied with some success to the registration of nuclear medicine images which are inherently noisy [Venot *et al.*, 1984; Venot and Leclerc, 1984; Venot *et al.*, 1983]. When the fluctuation level is low, the SSC criterion is no longer a similarity measure because the sign changes arising from mismatch of the object in the images may outnumber the statistical sign changes. For example, if two noiseless images are exactly matched, (1.13) predicts a zero *cost*, $S(a, b) = 0$. The zero *cost* is certainly the minimum in the displacement space because for any mismatch, $S(h, l) \geq 0$. For images which are less noisy, it is likely that the *cost* of match is lower than that of some mismatches of all possible translations. Therefore, the SSC criterion does not work well for images which are less noisy.

The DSC criterion overcomes the problem with the SSC criterion described above. It differs from the SSC criterion in the following two way:

1. f and g are assumed to have low noise levels;
2. Before generating the difference $d(i, j)$, the template f is modified such that

$$\begin{aligned} f'(i,j) &= f(i,j) + q, \qquad \text{if } i+j \text{ is even} \\ f'(i,j) &= f(i,j) - q, \qquad \text{if } i+j \text{ is odd.} \end{aligned} \tag{1.14}$$

   where q is a small real value.

Because of the low level noise assumption, the DSC criterion does not work well for images exhibiting large random fluctuations. This approach has been used to align high quality X-ray angiographic images of high signal to noise ratio (SNR) [Venot and Leclerc, 1984].

The Method of Differentials

Assume the difference between two images of an object captured at times t_1 and t_2 is entirely due to object motion. Therefore, the two images (denoted by $f_1(x, y)$ and $f_2(x, y)$) of the object are related by the coordinate transformation (1.5). So

$$f_2(x', y') = f_1(x, y) \tag{1.15}$$

where the relation between (x, y) and (x', y') is given by (1.5). Further assume that the translation and rotation are small. This gives

$$\cos\psi \simeq 1, \quad \sin\psi \simeq \psi \tag{1.16}$$

Therefore, from (1.5) and remembering that both ψ and a, b are small,

$$\begin{aligned} x' - x &\simeq y\psi - a - b\psi \simeq y\psi - a \\ y' - y &\simeq -x\psi + a\psi - b \simeq -x\psi - b \end{aligned} \tag{1.17}$$

It follows that, by expanding LHS (1.15) in a Taylor's series with respect to (x, y), one obtains

$$f_2(x', y') = f_2(x, y) + \frac{\partial f_2}{\partial x}(x' - x) + \frac{\partial f_2}{\partial y}(y' - y) + R \tag{1.18}$$

Substituting (1.15), (1.17) into (1.18) and ignoring the remainder R, one obtains the basic equation for calculating ψ and a, b (i.e. the rotation and the translation):

$$\Delta I(x, y) = \frac{\partial f_2}{\partial x}a + \frac{\partial f_2}{\partial y}b - (y\frac{\partial f_2}{\partial x} - x\frac{\partial f_2}{\partial y})\psi \tag{1.19}$$

where

$$\Delta I(x, y) = f_2(x, y) - f_1(x, y) \tag{1.20}$$

Equation (1.19) is linear in the three unknowns a, b and ψ. Therefore, if one chooses three points (or pixels in practice) and calculate the corresponding ΔI's and gradients, the rotation and translation can be uniquely determined. In particular, when there is only translation, the last term in (1.19) vanishes and only two points are needed for calculating the translation. Cafforio and Rocca [1983] discuss this topic in great detail.

Chapter 2

Spectral Motion Analysis

2.1 Introduction

Spectral motion analysis involves the extraction of motion parameters of objects, based on the use of Fourier transform [Bates and McDonnell, 1986, §6] properties of perceived images.

Fourier analysis is conventionally used to completely describe linear systems, an alternative to linear differential equations [Gabel and Roberts, 1980]. However, it has now been realised that it is an alternative representation of a wide range of physical quantities. The Fourier transform mathematically, as well as practically, provides complete information about an image in a domain (Fourier space) differing from the conventional spatial domain (image space). A unique, one-to-one relationship exists between a real world image and its Fourier transform, i.e. both are equally representative of the object. Therefore, the Fourier transform of an image should contain the same amount of information that the image contains. It is also well known that through the use of a mathematical transformation, one can often greatly simplify a problem, or solve a problem which may be impossible under its original representation.

In Chapter 1, the use of FFT technique for increasing the speed of computation for the matching algorithm is discussed. However, it is only a computational scheme in that context and should not be confused with the concept of spectral analysis mentioned above. On the other hand, due to the increasing importance of digital computers, the discrete representation of an image, and equivalently its Fourier transform, is essential for any practical computational process. Consequently, the discrete Fourier transform (DFT) is used to approximate the continuous case. It is understood that, mathematically, a continuous function and its samples are related by the Whittaker-Shannon Theorem [Goodman, 1968], i.e. the continuous function can be completely reconstructed from its samples, provided that the sampling frequency is equal to, or greater than, the Nyquist rate. The use of the FFT is implied whenever the computation of the Fourier transform is invoked in this thesis, and the size of each image, and of its Fourier transform, is required to be a power of 2. This ensures high computational efficiency (see §1.3.2).

The Fourier transform and its various properties which are important to motion analysis are introduced in §2.2. The computation of Fourier transforms by the DFT is also briefly discussed in §2.2. Some of the more widely employed techniques for analysing translational and rotational motion of an object are discussed in §2.3.

2.2 Fourier Analysis

2.2.1 The Fourier Transform

The Fourier transform of f, a function of two-dimensional image space vector $\mathbf{x}$, where f can in general be complex, is defined by [Bates and McDonnell, 1986]

$$F(\mathbf{u}) = \mathcal{F}\{f(\mathbf{x})\} = \int\int_{\infty}^{\infty} f(\mathbf{x}) \exp[i2\pi \mathbf{u} \cdot \mathbf{x}]\, dx\, dy \tag{2.1}$$

where $\mathbf{u}$ is the Fourier space vector representing the spatial frequency, $i = \sqrt{-1}$ and "$\cdot$" represents the scalar product. The inverse Fourier transform from which the complete image is recovered is given by

$$f(\mathbf{x}) = \mathcal{F}^{-1}\{F(\mathbf{u}\} = \int\int_{\infty}^{\infty} F(\mathbf{u}) \exp[-i2\pi \mathbf{u} \cdot \mathbf{x}]\, du\, dv \tag{2.2}$$

It is often convenient to express the Fourier transform in polar coordinates. Let $f(\mathbf{x}) = f(x,y) = f(r;\theta)$ and $F(\mathbf{u}) = F(u,v) = F(\rho;\phi)$, where (x,y) and (u,v) are Cartesian coordinates of $\mathbf{x}$ and $\mathbf{u}$ respectively, with $(r;\theta)$ and $(\rho;\phi)$ being the corresponding polar coordinates. It follows that

$$F(\rho;\phi) = \int_0^{\infty} r\, dr \int_0^{2\pi} d\theta\, f(r;\theta) \exp[i2\pi\rho r cos(\phi - \theta)] \tag{2.3}$$

Similarly

$$f(r;\theta) = \int_0^{\infty} \rho\, d\rho \int_0^{2\pi} d\phi\, F(\rho;\phi) \exp[-i2\pi\rho r cos(\phi - \theta)] \tag{2.4}$$

Because of the one-to-one relationship between the forward and inverse Fourier transform, it is convenient to use a notation which implies such an exclusive relation. The symbol $\longleftrightarrow$ is invoked to denote any Fourier transform pair. For example, $f(\mathbf{x}) \longleftrightarrow F(\mathbf{u})$, or equivalently $F(\mathbf{u}) \longleftrightarrow f(\mathbf{x})$ implies that $f(\mathbf{x})$ and $F(\mathbf{u})$ are Fourier transforms of each other.

The function $f(\mathbf{x})$ should satisfy two conditions [Bracewell, 1978; Goodman, 1968] for its Fourier transform to exist. The conditions are: 1) the integral of $|f(\mathbf{x})|$ from $-\infty$ to $+\infty$ exists; 2) if there are discontinuities in $f(\mathbf{x})$, there must be a finite number of them within any finite interval or region. Fortunately, these conditions do not exclude any real-world image of practical interest. Note that the Dirac δ-function, which is very useful for mathematical analysis, satisfies these conditions even though it has infinite value at a point and is discontinuous. The δ-function can

be conveniently defined as the limit of a *sinc* function [Bates and McDonnell, 1986, §6], i.e.

$$\begin{aligned}\delta(\mathbf{x}) &= \iint_{-\infty}^{\infty} \exp(i2\pi\mathbf{u}\cdot\mathbf{x})\,du\,dv \\ &= \lim_{T\to\infty}\iint_{-T/2}^{T/2} \exp(i2\pi\mathbf{u}\cdot\mathbf{x})\,du\,dv \\ &= \lim_{T\to\infty} T^2\mathrm{sinc}(Tx)\mathrm{sinc}(Ty) \qquad (2.5)\end{aligned}$$

Various useful expressions for the δ-function and related formulas can be found elsewhere [Bates and McDonnell, 1986; Goodman, 1968; Macovski, 1983]. It is straightforward to show that $\delta(\mathbf{x}) \longleftrightarrow 1$.

Some basic Fourier transform relationships [Bracewell, 1978; Goodman, 1968] are collected below.

Linearity: The Fourier transform operation is linear, so that the transform of the weighted sum of two functions is the weighted sum of their individual transforms:

$$a\,f(\mathbf{x}) + b\,g(\mathbf{x}) \longleftrightarrow a\,F(\mathbf{u}) + b\,G(\mathbf{u}) \qquad (2.6)$$

Magnification: The stretching of coordinates in one domain results in a proportional contraction in the other domain, together with a scaling of the transformation:

$$f(ax, by) \longleftrightarrow F(u/a, v/b)/|ab| \qquad (2.7)$$

Shift: Translation of a function in image space introduces a linear phase shift in frequency space. This is also referred to as the *shift theorem* [Bracewell, 1978]:

$$f(x-a, y-b) \longleftrightarrow F(u,v)\exp[i2\pi(ua+vb)] \qquad (2.8)$$

Convolution: The *convolution theorem* [Bates and McDonnell, 1986, §7] states that the convolution of two functions corresponds to multiplication of their transforms:

$$\iint f(\xi,\eta)\,h(x-\xi, y-\eta)\,d\xi d\eta \longleftrightarrow F(u,v)H(u,v) \qquad (2.9)$$

Cross Correlation: The *correlation theorem* [Bates and McDonnell, 1986, §7] states that the correlation of two functions also corresponds to multiplication of their transforms, but one of them is inverted:

$$\iint f(\xi,\eta)\,h(x+\xi, y+\eta)\,d\xi d\eta \longleftrightarrow F(-u,-v)H(u,v) \qquad (2.10)$$

In particular, when $h(x,y) = f(x,y)$ and both are real functions, (2.10) results in the widely applied autocorrelation theorem [Bates and McDonnell, 1986]

$$\iint f(\xi,\eta)\,f(x+\xi, y+\eta)\,d\xi d\eta \longleftrightarrow F^*(u,v)F(u,v) = |F(u,v)|^2 \qquad (2.11)$$

2.2.2 Evaluating Fourier Transforms via the DFT

In practice, every image is digitised so that its values are known at the nodal points of a finite rectangular mesh. Also, finite discrete transformations are used instead of continuous ones. The discrete Fourier transform (DFT) assumes that the function to be transformed is periodic. The finite extent of the function is treated as a single cell of an infinitely periodic function. In order that the DFT can be applied satisfactorily, the image and as many as possible of its derivatives must be repeated at points on opposite sides of a cell. However, most images obtained by practical means (e.g. from a TV camera) violate this requirement in one way or another. A common situation is that in which the acquired data is only part of a scene, i.e. a truncated image is obtained. Equivalently, the truncated image can be seen as the result of multiplying the whole scene by a rectangular window. In the Fourier domain, this corresponds to convolving the true spectrum with a *sinc* function. The truncation prevents the true spectrum from being estimated exactly. The truncation error is often termed *spectral leakage* [Harris, 1978].

There are two proven techniques for reducing spectral leakage. They are windowing [Harris, 1978; Childers, 1978] and edge-extension [Bates and McDonnell, 1986, §15]. By windowing a quantity, one forces it to fall to zero gradually at its extremities. The shape of the window is important in practice. The art of designing windows has been called *window carpentry* by Yuen and Fraser [1979]. The edge-extention technique, which was developed at the University of Canterbury by Bates and McDonnell [1986] [cf. McDonnell; 1975], involves extending the image at its edges in such a way that the extended image and its first derivatives (and, sometimes, other derivatives) are continuous at the boundary of each periodic cell.

2.2.3 Shift-Rotation Invariant Properties of the Fourier Transform

In order to study the effect of motion on the object's Fourier transform, it is important to know how the spectrum is modified by the spatial changes associated with motion. Two particular features of the Fourier transform relating directly to motion estimation are its translational and rotational *invariant properties* [McCallum *et al.*, 1986; Gardenier *et al.*, 1986]. Recalling the coordinate transformation discussed in §1.3.1, the image of a shifted object is denoted as $g(x,y) = f(x',y')$. The Fourier transform of $g(x,y)$ is seen, from (2.1), to be

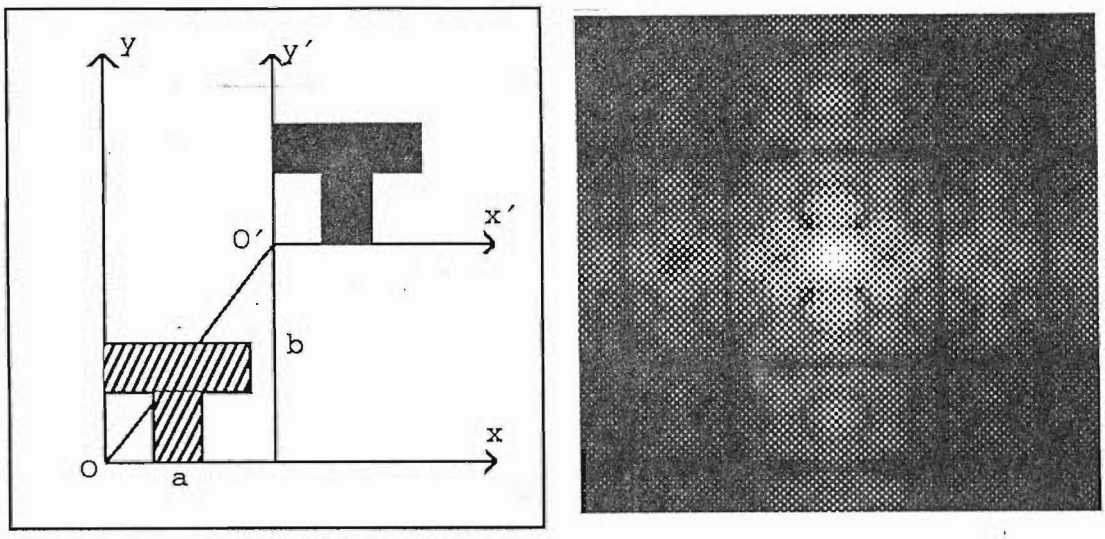
$$G(u,v) = \int\!\!\int_{-\infty}^{\infty} f(x',y') \exp[i2\pi(ux+vy)]\, dx\, dy \tag{2.12}$$

From (1.2) and the shift theorem (2.8), one sees for a translation (a,b) that

$$G(u,v) = F(u,v) \exp[i2\pi(ua+vb)] \tag{2.13}$$

Taking the modulus of both sides of (2.13), one obtains the following *shift-invariant* property of the Fourier magnitude:

$$|G(u,v)| = |F(u,v)| \tag{2.14}$$



(a) (b)

Figure 2.1. The shift-invariant property of Fourier transform. (a) The object "T" translates by a shift vector (a, b). The hatched "T" indicates the object's original position, and the black "T" represents the final position. (b) the Fourier transform magnitude of the "T" remains unchanged after the translation. Note that although the object "T" is black in (a), the object and the background intensities are treated, respectively, as 1 and 0 for computing the Fourier transform.

which also indicates that all information regarding the translational shift of an object is contained in the difference of the phases of $G(u, v)$ and $F(u, v)$. This invariant property is illustrated in Fig. 2.1. Note that although the object "T" is black in Fig. 2.1(a), the object and the background intensities are treated, respectively, as 1 and 0 for computing the Fourier transform. This rule applies to all two-valued (including hard-limited) images throughout the rest of this thesis.

Now consider rotation through an angle ψ, as illustrated in §1.3.1. By substituting (1.3) into (2.12), with some rearrangement and remembering that $dx'dy' = dxdy$, it is seen that

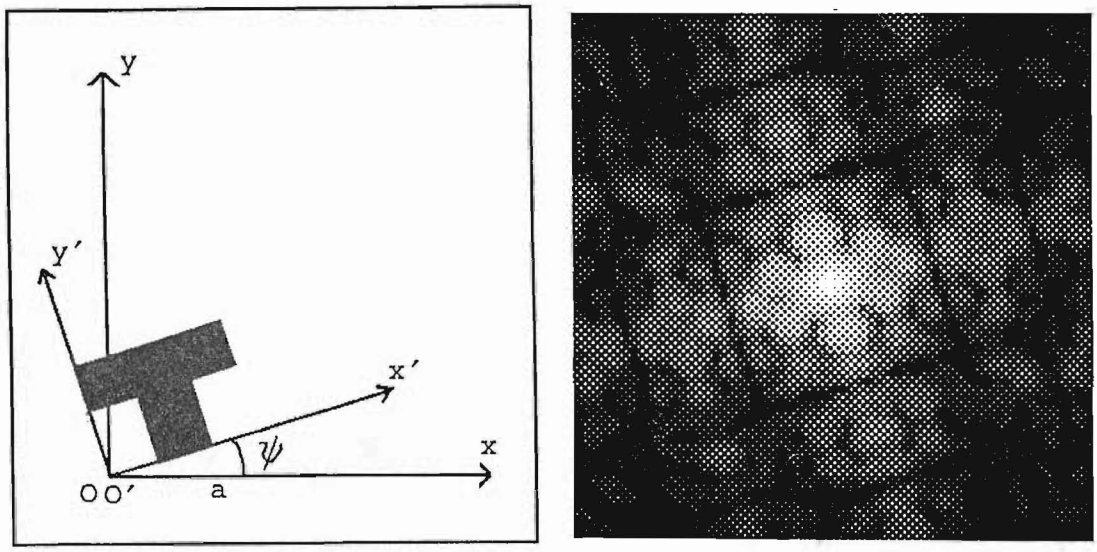
$$G(u, v) = F(u', v') = F(u\cos\psi + v\sin\psi, -u\sin\psi + v\cos\psi) \tag{2.15}$$

This gives

$$\begin{aligned} u' &= u\cos\psi + v\sin\psi, \\ v' &= -u\sin\psi + v\cos\psi \end{aligned} \tag{2.16}$$

Therefore, on referring to (1.3), one obtains the Fourier space coordinate transformation relations corresponding to rotation in image space:

$$\begin{aligned} u &= u'\cos\psi - v'\sin\psi, \\ v &= u'\sin\psi + v'\cos\psi \end{aligned} \tag{2.17}$$



(a) (b)

Figure 2.2. The rotational invariant property of Fourier transform. (a) The object "T" rotates through an angle ψ (compare with Fig. 2.1(a)). (b) the Fourier transform magnitude of the rotated "T" after the rotation.

which is formally identical to (1.3), indicating that the corresponding change in Fourier space is the same as that in image space, i.e. a rotation of ψ in this case. This is referred to here as the *rotational invariant property.* Fig. 2.2 illustrates this property.

It is often more convenient to discuss rotation in terms of polar coordinates. When there is only rotation, it is seen from (2.3) that

$$G(\rho;\phi) = \int\!\!\int_{-\infty}^{\infty} f(r;\theta+\psi)\exp[i2\pi\rho r cos(\phi-\theta)]\, r\, d\theta\, dr = F(\rho;\phi+\psi) \qquad (2.18)$$

In general, object motion may involve both translation and rotation. Because these two operations do not commute (see Fig. 2.3) as discussed in §1.3.1, their order affects the final outcome. It turns out that the two corresponding Fourier transforms differ from each other in the linear phase terms caused by the translation. For the situation of translation followed by rotation (see Fig. 2.3(a)), it is seen from (2.12) and (1.6), respectively, that

$$G(u,v) = F(u',v')\exp[i2\pi(u'a+v'b)] \qquad (2.19)$$

where u', v' is given by (2.16). For the case of rotation followed by translation (see Fig. 2.3(b)), one has from (1.5) that

$$G(u,v) = F(u',v')\exp[i2\pi(ua+vb)] \qquad (2.20)$$

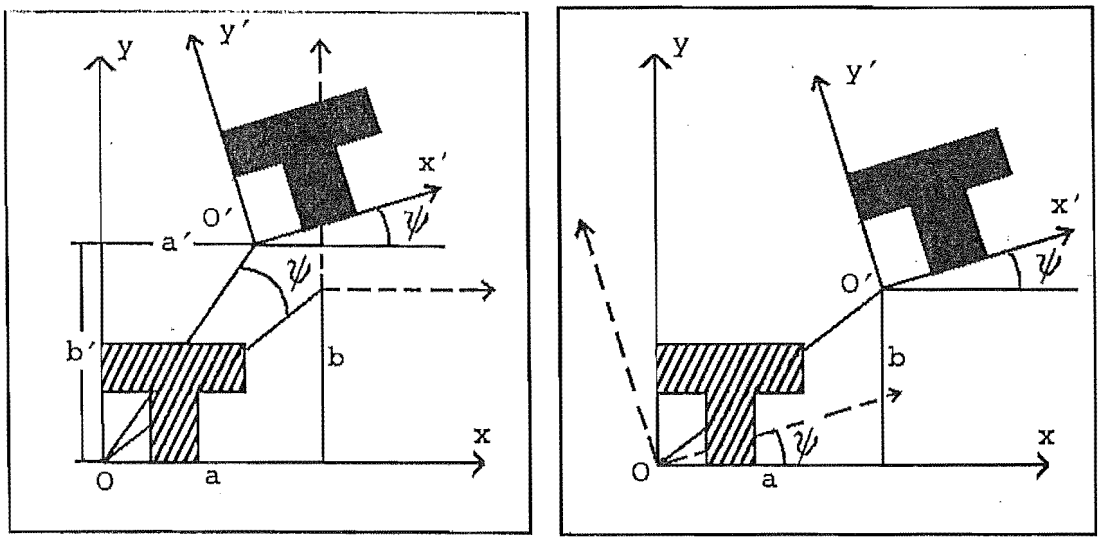


Figure 2.3. Illustration of the incommutiveness between the translation and the rotation. Both operations are taken with reference to the x, y coordinate system. (a) The object "T" translates by a shift vector (a, b), then rotates through an angle ψ. The final position and orientation of the object can be equivalently obtained by a rotation through an angle ψ, first then a translation of (a', b'). (b) The object "T" rotates through an angle ψ then translates by a shift vector (a, b). Note that, in both pictures, the hatched "T" shows the object's original position, the dashed axes indicates the intermediate step, and the black "T" represents the final position. The Fourier transform magnitude of the "T" is identical to both (a) and (b) situations, and is the same as shown in Fig. 2.2(b).

However, by referring to the argument on equivalent shifts in §1.3.1, one may rewrite (2.19) as

$$G(u, v) = F(u', v') \exp[i2\pi(ua' + vb')] \tag{2.21}$$

where a', b' are the equivalent shifts introduced in (1.7). Taking the modulus of both sides of (2.20) and (2.21), one sees that $|G(u, v)| = |F(u', v')|$. This means that the rotational invariant property of the Fourier magnitude is preserved in both situations (see Fig. 2.2(b)).

In summary, with translation alone, the Fourier transform magnitude is unchanged and the phase has an extra linear term. With rotation, the Fourier transform rotates through the same angle as the image. When there is both rotation and translation, the Fourier transform magnitude is preserved, but it is rotated.

2.3 Motion Estimation

2.3.1 General Remarks

It is clear from the last section that it is possible to separate the effects of translation and rotation in image space by considering the phase and magnitude separately in Fourier space. This is an important advantage of Fourier domain techniques over image domain techniques. Another advantage of Fourier techniques is that the ability to deal with rotation independently dramatically speeds up the estimation of both translational and rotational motion. Fourier techniques are also the basis of optical processing or optical computing [Casasent, 1985], especially in the area of pattern recognition [Casasent and Psaltis, 1977; Casasent, 1981].

Fourier mapping is a global operation. The separation or extraction of features of multiple objects is often impracticable in the Fourier domain. Fourier methods cannot usually be applied directly to the types of multiple-object motion discussed in §1.2.1. It is of course always possible to attempt to recognise and separate the part of an image containing a particular object from the parts containing other objects, and then apply Fourier techniques to the separated part. Particularly, one can also combines the use of Fourier methods with that of the spatial domain techniques in the manner that is described by Goshtasby [1985] on the template matching employing moments (see §1.2.3). However, one can sometimes apply segmentation techniques in Fourier space. Invoking the terminology of optical science, the Fourier transform of an image containing several objects manifests the interference which coherent light from sources representing the various objects waves would exhibit in the focal plane of a lens. Fourier transforms of images exhibiting pronounced detail tend to be strongly peaked [Casasent, 1985] and the distinct peaks may be regarded as features in Fourier space. Little attention seems to have been paid in the literature to using the features in Fourier space directly for image segmentation.

If there is only translational motion, the most popular methods of estimation are based on using the phase difference of the Fourier transforms of two images of an object [Huang and Tsai, 1981; Kuglin and Hines, 1975]. It is straightforward to see from (2.13) that this phase difference is a linear function of the object shift. This idea is attractive since it is conceptually simple. The major computational task is calculating the Fourier transform. The magnitude of the Fourier transform of the difference between a pair of images of the object also fully characterises the change in the object's position in one of the images relative to the other. This possibility does not seem to have been described before. However, my study shows that a complicated search algorithm would need to be invoked in addition to Fourier transformation. An image's cepstrum [Bogert *et al.*, 1963; Oppenheim and Schafer, 1975] can also be invoked for detecting translation, as described by Lee *et al.* [1988] and Mitra *et al.* [1988]. Methods of estimating translation through phase difference approaches are discussed in detail in §2.3.4. Other two approaches (i.e. magnitude of the Fourier transform of the difference between two images, and cepstrum) are introduced in §2.3.5.

When there is only rotation, one can simply match features of the Fourier trans-

form magnitudes, as described in §2.3.6. An improved method (also described in §2.3.6) which matches corresponding features of the Fourier transform magnitudes together with their phases tends to give more accurate results. It should be noted that one can certainly match the entire contents of the two Fourier transforms, or Fourier transform magnitudes, without subjectively selecting features. However, this dramatically increases the computational cost.

In situations where the object both rotates and translates, a general approach is to estimate the rotation (from the Fourier magnitude) and then, after compensating for rotation, estimate the translation.

In addition to rotation and translation, scale change or magnification (such as what happens to the image when the object moves away from or towards the camera) can also be treated separately in the spectral domain [Huang and Tsai, 1981]. For example, one can match two translated objects' Fourier magnitudes by varying the scaling factor. The image and its Fourier transform scale in opposite senses. The Mellin transform [Casasent and Psaltis, 1977] is scale invariant and it can be applied to images to remove the scale differences before estimating rotation and translation.

This thesis also proposes a new method for estimating rotation. The method employs the third order spectrum, or bispectrum, or Erlangen bispectrum [Bates, 1987] rather than the ordinary spectrum (Fourier transform). It is still, by its nature, a Fourier domain approach (see Chapter 3). However, the Erlangen bispectrum has distinct characteristics differing from those of the ordinary spectrum. The bispectrum and its application to motion estimation are presented in Chapter 3 and Chapter 5 respectively.

2.3.2 Resampling and Interpolation

Computations associated with motion estimation involve resampling, for two reasons. Firstly, data samples are usually available at the points of intersection of a rectangular (usually square) mesh, consistent with a Cartesian coordinate system. However, data at points other than these are generally needed for estimating rotational motion. Secondly, in translational motion detection it is often desirable to estimate the translation to a higher resolution then the unit distance between pixels (i.e. samples) [Tian and Huhns, 1986]. The term "subpixel registration" is sometimes used to describe such an algorithm. Thus most motion estimation algorithms require resampling procedures, most conveniently effected by various interpolation techniques [Bates and McDonnell, 1986].

A near-perfect interpolation may be achieved for band-limited data which has been uniformly sampled on a square mesh (assuming samples are sufficiently closely spaced) by the use of appropriate *sinc* functions [Bates and McDonnell, 1986, §11]. However, the computational burden of *sinc* interpolation limits its practical use since each interpolating function in each coordinate has non-zero values across the extent of that coordinate within the image. Therefore a short-range interpolation scheme, such as nearest neighbour, linear, Lagrange polynomial, or cubic spline [Bates and McDonnell 1986, §§11, 33 & 47] is often invoked. Another scheme, which is preferred

for the approaches to estimating rotational motion described in this thesis, is *bilinear interpolation* [Bates and McDonnell, 1986, §45][cf. Rosenfeld and Kak 1982]. This scheme is implemented as follows: given a set of samples $f_{i,j}$ at points $\mathbf{x}_{i,j}$ on a rectangular mesh, linear interpolation is first performed along grid lines in one direction, e.g.

$$\tilde{f}(x, y_j) = f(x_i, y_j) + (f(x_{i+1}, y_j) - f(x_i, y_j))\frac{x - x_i}{\Delta x} \quad x_i \leq x < x_{i+1} \tag{2.22}$$

where Δx is the distance between samples in the x direction. Then linear interpolation is performed in the perpendicular direction, so that

$$\tilde{f}(x, y) = \tilde{f}(x, y_j) + (\tilde{f}(x, y_{j+1}) - \tilde{f}(x, y_j))\frac{y - y_j}{\Delta y} \quad y_j \leq y < y_{j+1} \tag{2.23}$$

It can be seen that at any interpolated point only the 4 nearest neighbours are used, so that bilinear interpolation is computationally efficient.

2.3.3 Phase-Only Approaches and Phase Dominance

During the past 20 years, in acoustical, optical and digital image processing, Fourier techniques which make use of only transform phase have found important applications. The use of phase-only information has been studied in many different fields. For instance, Oppenheim and Lim [1981] and Bates and Fright [1985] emphasise the overiding importance of phase in general, as recognised by Metherell (1969) for holography and Ramachandran and Srinivasan [1970] for X-ray crystallography. Optical holographers [cf. Lesem *et al.* 1969] have also constructed phase-only computer-generated holograms called kinoforms. Other workers [cf. Powers *et al.* 1970; Lohmann 1970; Oppenheim and Lim 1981; Oppenheim *et al.* 1983; Anderson and Anderson 1986] have shown in a variety of acoustical and optical contexts that phase-only holograms reconstruct recognisable images, while amplitude-only holograms do not.

Horner and Gianino [1984] have proposed the *phase-only matched filter* or *phase-only correlation filter* technique for optical processing. Their studies, supplemented by others [Gianino and Horner, 1984; Anderson and Anderson, 1987] have shown the superior performance of such filters as compared with that of the classical matched filter [VanderLugt, 1964], especially as regards the sharpness of the correlation peak, discrimination capability for nearly similar objects, and sensitivity to scaling and rotational effects. This technique has found interesting application in optical pattern recognition [Davis *et al.*, 1987; Cottrell *et al.*, 1987].

The *phase correlation* (PC) method [Kuglin and Hines, 1975], a phase difference technique for motion estimation discussed in §2.3.4, has found increasing application, especially in the context of digital image processing. Video rate phase correlation has been implemented in hardware for processing image sequences [Pearson *et al.*, 1977]. Other researchers [Kuglin and Hines, 1975; Pearson *et al.*, 1977; Alliney and Morandi, 1986; Castro *et al.*, 1987; Wu *et al.*, 1989] have pointed out the superior performance and properties of this technique over ordinary cross-correlation,

and have applied it to solve medical [Wu *et al.*, 1989; Castro *et al.*, 1987] and aircraft guidance [Kuglin *et al.*, 1979] image registration problems. It has also been applied for coding of image sequences by Lohamnn and Thum [1983].

In §2.3.4, the concept of phase dominance is invoked to interpret the performance of phase correlation. This idea is also invoked in Chapter 4 when explaining the efficacy of windowing in phase correlation.

2.3.4 Estimating Translational Motion — Phase Difference Approaches

Let $f(x,y)$ and $g(x,y)$ denote two images of an object. The two images are recorded while the object translates. $F(u,v) \longleftrightarrow f(x,y)$ and $G(u,v) \longleftrightarrow g(x,y)$ are the corresponding Fourier transforms.

Huang and Tsai's Method

Huang and Tsai [1981] have suggested a method of estimating the translation on the basis of calculating the phase difference between $G(u,v)$ and $F(u,v)$ at discrete points. From (2.13), it is seen that

$$\Delta\Phi(u,v) = \text{Phase}\{G(u,v)\} - \text{Phase}\{F(u,v)\} = 2\pi(ua + vb) \tag{2.24}$$

Therefore, if one calculates $\Delta\Phi(u,v)$ for a pair of spatial frequencies, say (u_1,v_1) and (u_2,v_2), one can solve for a and b from a pair of simultaneous equations.

Equation (2.24) indicates that before calculating the phase difference $\Delta\Phi(u,v)$, $G(u,v)$ and $F(u,v)$ have to be computed. In order to simplify the computation, one can compute

$$\Delta\Phi(u,0) = \text{Phase}\{G(u,0)\} - \text{Phase}\{F(u,0)\} = 2\pi ua \tag{2.25}$$

so that a can be estimated directly. Likewise, b can be obtained from

$$\Delta\Phi(0,v) = \text{Phase}\{G(0,v)\} - \text{Phase}\{F(0,v)\} = 2\pi vb \tag{2.26}$$

However, care must be exercised when calculating the phase differences in (2.25) and (2.26) since only the principal value (i.e. in the range of $-\pi$ to $+\pi$) of Phase$\{F(u,v)\}$ is known. If the principal values of Phase$\{G(u,0)\}$ and Phase$\{F(u,0)\}$ are used in (2.25) and (2.26), the resulting $\Delta\Phi(u,0)$ may not be correct. Huang and Tsai [1981] have suggested a method to overcome this problem. While their method is computationally inexpensive, it is noise-sensitive like other methods which are based on estimation at discrete points. It is therefore desirable to repeat the calculation throughout Fourier space and invoke least-squares estimation.

Phase Correlation

The phase-correlation (PC) technique, which is one of the major concerns of this thesis, can be formulated directly from (2.13). The quantity $D(u,v)$

$$D(u,v) = \frac{G(u,v)}{F(u,v)} = \exp[i2\pi(ua+vb)] \tag{2.27}$$

is called the phase-difference spectrum. Phase correlation is achieved by computing the inverse Fourier transform of $D(u,v)$, i.e.

$$\mathcal{F}^{-1}\{D(u,v)\} = \iint \exp[i2\pi(ua+vb)]\exp[-i2\pi(ux+vy)]\,du\,dv = \delta(x-a,y-b) \tag{2.28}$$

It is seen that the inverse Fourier transform results in a δ-function. Therefore, by recognising the position of the δ-function in the phase correlation, the translation is found. In practice, of course, the δ-function becomes a peak of finite width. The consequences of this are discussed in detail later in this subsection.

Although in principle simple, the division in (2.27) is sensitive to distortions such as noise. This can be explained by the division process. When the denominator $F(u,v)$ is very small, for example, for high spatial frequencies, a deviation from the true value due to the distortions can be reinforced by the division.

Previous workers [Kuglin and Hines, 1975; Alliney and Morandi, 1986; Castro *et al.*, 1987] compute the phase-difference spectrum in the following way.

$$D(u,v) = \frac{G(u,v)F^*(u,v)}{|G(u,v)F^*(u,v)|} = \exp[i2\pi(ua+vb)] \tag{2.29}$$

This avoids the difficulties associated with division (when $F(u,v)$ tends to 0, $D(u,v)$ tends to 1). Equivalently, an alternative way of implementing phase correlation, as proposed by Wu *et al.* [1989], combines the phase difference $\Delta\Phi(u,v)$, as defined in (2.24), with unity magnitude to form the phase-difference spectrum, i.e.

$$D(u,v) = \exp[\Delta\Phi(u,v)] = \exp[i2\pi(ua+vb)] \tag{2.30}$$

An extension to this implementation is described in §4.2.

Alliney and Morandi [1986] have considered invoking the projection theorem [Bates and McDonnell, 1986, §9] for reducing the overall computation of phase correlation. It is convenient to compute the two projections parallel, respectively, to the x-axis and to the y-axis. These two projections are here referred to as the x and y projections. The x projections of the two images $f(x,y)$ and $g(x,y)$ are, by definition [Bates and McDonnell, 1986, §9],

$$p_x^f = \int f(x,y)\,dy \quad \text{and} \quad p_x^g = \int g(x,y)\,dy \tag{2.31}$$

It follows from the projection theorem, that

$$p_x^f \longleftrightarrow F(u,0) \quad \text{and} \quad p_x^g \longleftrightarrow G(u,0) \tag{2.32}$$

By referring to (2.25) and (2.30), a 1-D phase difference spectrum $D(u)$ can be constructed as

$$D(u) = D(u,0) = \exp(i2\pi ua) \tag{2.33}$$

Therefore, $\mathcal{F}^{-1}\{D(u)\}$ results in a 1-D phase correlation from which the shift a can be estimated. In the same manner, b can be obtained. Since only 1-D Fourier transformations are involved, the computational load is light.

Comparison of the Phase and the Cross Correlation Techniques

It is instructive to compare PC with the most widely invoked cross correlation (CC) technique when discussing the merits and demerits of the PC technique. The phase correlation technique is less demanding than the cross correlation technique as regards both computer time and storage. In order to calculate the cross correlation coefficients using the FFT algorithm, one needs to extend the size of the image array by zero packing because the content of the cross correlation is usually very broad. Suppose one matches two $N \times N$ pixel images. The image array containing the cross correlation should be $2N \times 2N$, so that $4N^2 \log_2(4N^2)$ multiplications are needed. If a smaller image array is used, errors occur due to aliasing in the correlation domain [Gonzales and Wintz, 1977, §4]. However, this is not the case for phase correlation, because the magnitude of the phase difference spectrum is unity, which ensures that the content of its inverse Fourier transform (see, for example, Fig. 4.5) is very narrow (ideally, $\mathcal{F}^{-1}\{D(u,v)\}$ is a centre-shifted $\delta-$function). Therefore, to calculate the Fourier transform of a $N \times N$ image one needs only $N^2 \log_2 N^2$ multiplications to implement the phase correlation method. Hence, it is seen that the ratio of computation time for CC to that for PC is given by

$$\frac{4N^2 \log_2(4N^2)}{N^2 \log_2 N^2} = 4 + \frac{4 \log_2 4}{\log_2 N^2} \tag{2.34}$$

implying that the PC method is at least 4 times as efficient computationally as the CC algorithm for the example given above.

Real-world images always suffer from various distortions and constraints, as discussed in §1.2.1. These imperfections degrade the performance of phase correlation, so that one ends up with only an approximation to a δ-function. However, most of the energy in the phase correlation is usually concentrated at the position where, ideally, the peak of the δ-function should be. In general, the broader a spectrum's extent, the narrower the corresponding image's extent [Bracewell, 1978]. Therefore, an advantage of PC is that the registration peak tends to be appreciably sharper than for CC [Gianino and Horner, 1984; Anderson and Anderson, 1987; Castro *et al.*, 1987; Wu *et al.*, 1989] and, as a result, PC usually exhibits better accuracy than CC.

It is also worth noting that the characteristic of PC discussed in the last paragraph is similar to that of the technique of matching edges or curves parts of an object described in §1.2.2 and §1.2.3. All such techniques can be further explained in terms of the concept of phase dominance (discussed in §2.3.1) as follows. Denoting the

gradient function of $f(x)$ as $f'(x)$, it is easy to verify that

$$f'(x) \longleftrightarrow -i2\pi u F(u) \tag{2.35}$$

Because $|F(u)|$ is usually small where u is large, $|uF(u)|$ is therefore a widened spectrum. The edges and curves in images are where intensities change quickly so that the corresponding gradients have large (absolute) values. Equation (2.35) indicates that high gradients contribute more to high frequencies, and therefore spectrum widening enhances the information of edges and curves. The phase-only spectrum is a kind of widened spectrum for which the magnitude is unity, and it can be used to reconstruct an edge and curve enhanced version of the original image [Oppenheim and Lim, 1981]. It could therefore be said that image features (edges and curves) are contained in the Fourier phase, which is of course just a manifestation of phase dominance.

Another advantage of PC over other methods of translation estimation, noted by Kuglin and Hines [1975], is that the technique is comparatively insensitive to variations in illumination (similar to the feature-based matching techniques described in §1.2.3). For example, for satellite images of the same scene taken at different time of day, the overall intensity distribution across each image usually varies [Svedlow *et al.*, 1978]. The intensity-based matching techniques described in §1.2.2 are likely to fail in this situation. From the viewpoint of matching, boundaries and curves are usually the most important attributes of objects. Since phases are related to these attributes rather than the absolute image intensity distribution, as described in the last paragraph, the PC technique is relatively insensitive to variations in illumination.

Unlike matching by the CC or SAVD, where usually a small template is matched to a larger image as described in §1.2.2, the PC technique works on images of the same size. One source of error is the difference in image content between the two images near the edges, as illustrated in Fig. 2.4. The truncation effect, discussed in §2.2.2, also affects the performance of the PC technique. The use of windowing for reducing errors in phase correlation has been studied by Alliney and Morandi [1986] and is also discussed in §4.3.2 of this thesis in the context of implementation (see Fig. 4.8). However, straightforward implementations of phase correlation without windowing give surprisingly good results in many situations. This has been the experience of both myself and some other researchers [Alliney and Morandi, 1986; De Castro and Morandi, 1987] (see Fig. 4.5). The phenomenon has been considered by Alliney and Morandi [1986] when studying PC using only two projections. They demonstrated that straightforward implementations of phase correlation on 1-D projections do not give meaningful results even under the condition that the only error source is the content's difference at the edge. Their study, based on assuming that the edge difference is a stochastic process, shows that 2-D phase correlation is much more advantageous than the 1-D case. My feeling is that their result only partly explains this phenomenon. An alternative explanation, with illustrations, from the aspect of image feature content, is given in §4.3.2.

As a method of estimating translation, phase correlation is very sensitive to rotation. De Castro and Morandi [1987] show, in their study of image registration involving object rotation, that the PC technique can tolerate a rotation angle of $< 3°$. My studies are in broad agreement, although some images tested could tolerate rotations of less than 5° before translation detection by PC is inaccurate. The

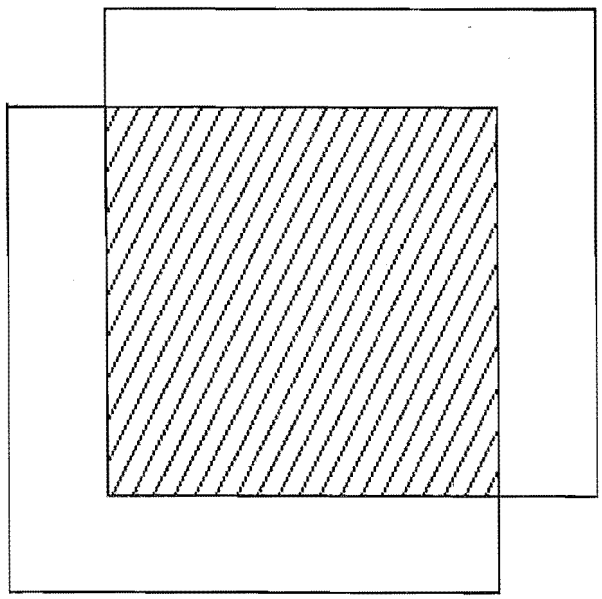


Figure 2.4. Illustration of the edge differences. The hatched region is common to both images.

limit of tolerance to rotation of the PC technique depends on the types of images being processed. However, it seems reasonable to conclude that this method can tolerate rotation of a few degrees. This problem is further studied and illustrated in Chapter 4.

Subpixel registration in PC improves the resolution of shift detection since the phase correlation peak can be better resolved. It has been reported that by invoking a quadratic interpolation at the region containing the peak and its four surrounding points, a resolution of 0.08 pixel can be achieved by the PC technique [Pearson *et al.*, 1977].

The PC technique cannot be applied to video image sequences obtained in many practical situations, as discussed in Chapter 4. A new form of phase correlation, which is termed the *sequential predictive phase correlation* (SPPC), is developed in that chapter. It is shown in Chapter 7 that the SPPC technique is a promising technique for estimating coronary artery motion from angiographic image sequences.

2.3.5 Estimating Translational Motion — Other Approaches

Fourier Magnitude of Difference Image

Denoting the difference of images $f(x,y)$ and $g(x,y)$ as $d(x,y)$, i.e.

$$d(x,y) = g(x,y) - f(x,y) \tag{2.36}$$

The Fourier transform of $d(x,y)$, denoted by $\mathcal{D}(u,v)$, is, from the linearity of the Fourier transform,

$$\mathcal{D}(u,v) = G(u,v) - F(u,v) = F(u,v)(\exp[i2\pi(ua+vb)] - 1) \tag{2.37}$$

The magnitude of $\mathcal{D}(u,v)$ is seen to be

$$|\mathcal{D}(u,v)| = |F(u,v)|\sqrt{2(1-\cos 2\pi(ua+vb))} \tag{2.38}$$

The zeros of $|\mathcal{D}(u,v)|$ appear at

$$2\pi(ua+vb) = 2n\pi, \qquad n = 0,1,2,.... \tag{2.39}$$

This corresponds to a family of straight lines of equal slope, i.e.

$$ua+vb = n \tag{2.40}$$

Therefore, one can estimate the translation (a and b) by detecting these straight parallel lines (which have zero intensity). In the presence of noise, one may need a complicated algorithm to search for these zero lines.

Power Cepstrum Approach

Very recently, Lee *et al.* [1988; cf. Mitra *et al.* 1988] have proposed employing the power cepstrum for image registration. The cepstrum technique was introduced by Bogert *et al.* [1963] for analysis of 1-D signals of time series containing echoes, for example, for signals of the form $s(t) = f(t) + w\, f(t-\tau)$, where w is a scale factor. The application of the cepstrum technique to 2-D data was studied by Dudgeon [1977]. To detect the translation between the input image $f(x,y)$ and the reference $g(x,y)$, the summation

$$l(x,y) = f(x,y) + h\, g(x,y) = f(x,y) + h\, f(x+a, y+b) \tag{2.41}$$

is first performed, where $0 \leq h \leq 1$ is the scale factor. The power cepstrum $\mathcal{P}\{f(x,y)\}$ of the image $f(x,y)$, is defined by

$$\mathcal{P}\{f(x,y)\} = |\mathcal{F}^{-1}\{\ln|F(u,v)|^2\}|^2 \tag{2.42}$$

Employing a series expansion, the power cepstrum of $l(x,y)$ is

$$\mathcal{P}\{l(x,y)\} = \mathcal{P}\{f(x,y)\} + A\delta(x,y) + B\delta(x\pm a, y\pm b) + C\delta(x\pm 2a, y\pm 2b) + ... \tag{2.43}$$

where A, B, C, etc are constant coefficients. This equation indicates that the total power cepstrum of $l(x,y)$ is the power cepstrum of $f(x,y)$ plus a train of δ-functions occurring at integer multiples of the translation. After subtracting $\mathrm{P}\{f(x,y)\}$ from $\mathrm{P}\{l(x,y)\}$, only the peaks (δ-functions) are left and the translation can therefore be estimated by detecting these peaks. When the scale factor h of (2.41) is chosen to be close to 1, the coefficients $A, B, C, ...$ in (2.43) are large, so that the peaks are easily detected [Lee *et al.*, 1988].

This technique was applied to the registration of fundus images (vascular structure of the eye) [Mitra *et al.*, 1988]. It is claimed by Lee *et al.* [1988] that the algorithm works fast and accurately compared to the conventional cross correlation technique.

2.3.6 Estimating Rotational Motion

It should be noted that not much attention has been paid in the literature to the practical implementation of rotational motion estimation in the Fourier domain, probably because most researchers have become incorrectly, in my opinion, convinced that the computational demands are excessive. Only a few researchers [Huang and Tsai, 1981; McCallum *et al.*, 1986; De Castro and Morandi, 1987] have specifically considered spectral domain estimation of rotational motion.

Rotation may be estimated with the aid of Fourier transforms by straightforward application of the rotational invariant property described in §2.2.3.

Huang and Tsai [1981] describe the possibility of estimating relative rotation by comparing Fourier magnitudes of two images of an object. Most man-made objects contain sharp edges which give rise to distinct spectral domain lines through the origin in the direction orthogonal to the original edges in the image (see Fig. 2.1(b)). By superimposing the magnitudes and matching them by rotating one against the other, the angle of rotation can be estimated. Huang and Ysai suggest that a distinct line in one Fourier magnitude may be chosen to match the corresponding line in the other magnitude in order to reduce the computational cost.

McCallum *et al.* [1986] have developed an algorithm for automatically determining rotation angle in their study of template matching using Fourier descriptors [cf. Gardenier *et al* 1986]. The algorithm searches for a radius, along which the integral of the Fourier magnitude is a maximum, for a pattern to be matched. The orientation of the maximum integral radius is assumed to be known for the template. The relative rotation angle is therefore estimated by calculating the angle difference between the maximum integral radii for the pattern and the template.

De Castro and Morandi [1987] have considered combining the estimation of rotation with that of translation by means of phase correlation. The basic idea is that, after rotation is corrected, the PC should ideally result in a δ-function. The procedure of estimation is as follows. Rotate one Fourier transform against the other. For each rotated angle compute the corresponding PC. Compare those resulting phase correlations. The one which is the closest to an ideal δ-function among the many is chosen, and the position of its peak determines the translation. The rotation angle corresponding to the chosen phase correlation is just the estimate of rotation. The computational load is very heavy since there are a great many of PCs. In order to reduce the number of PCs computed, De Castro and Morandi [1987] suggest that PCs be computed only at angular intervals around angles which are suspected to be close to the rotation angle of the object.

A general Fourier domain matching algorithm for estimating rotation can be devised based on the *angular correlation* defined in what follows. For convenience, polar coordinates are adopted in Fourier space. In the case of motion with both translation and rotation, the angular correlation of $|F(\rho;\phi)|$ and $|G(\rho;\phi)|$ is given by

$$\Gamma(|F|,|G|,\rho;\phi) = \left| \int_0^{2\pi} |F(\rho;\phi')||G(\rho;\phi'+\phi)|\, d\phi' \right| \tag{2.44}$$

As indicated in §1.2.2, correlation techniques possess the advantage of suppressing

noise. This is the major motivation for invoking correlation techniques for estimating rotation.

Although Fourier magnitude sufficiently determines the object rotation by means of angular correlation, the inability to use phase information makes the method sensitive to distortions as described later in Chapter 5. The angular correlation with phase is here defined by

$$\Gamma(F, G, \rho; \phi) = \left| \int_0^{2\pi} F(\rho; \phi') G^*(\rho; \phi' + \phi)\, d\phi' \right| \tag{2.45}$$

It is seen to have the form of complex correlation. In Chapter 5, it is also shown that the estimation algorithm based on (2.45) is much more robust and gives higher accuracy then the magnitude only angular correlation. The point is that the angular correlation of the magnitude only method is smooth everywhere and the correlation peak is broad. With the additional phase information, the correlation values are mostly eliminated except at the position of the correlation peak, which is therefore much narrower then that for the magnitude only correlation.

The computation of angular correlation described here involves two steps. The first is interpolation to resample data at the required polar coordinates. The second step is the correlation integration. In order to reduce the overall computational cost, the integrations based on (2.44) and (2.45) can be made with respect to a selected ρ so that only the data on a circular ring of a radius ρ is employed for computing the motion.

Chapter 3

Bispectral Analysis

3.1 Introduction

The bispectrum (the Fourier transform of the triple correlation) [Bates, 1987] was introduced into signal/image processing [Huber *et al.*, 1971] some 20 years ago, although it had been discussed earlier in the context of statistics [Akaike, 1966; Brillinger, 1965]. It has received significant attention from electrical engineers and applied physicists only within the last 10 years [Lohmann and Wirnitzer, 1984; Nikias and Raghuveer, 1987]. Bispectral estimation has recently been considered in a digital signal processing framework by Nikias and Raghuveer [1987]. It is known that bispectral estimation, which is useful for various tasks [Nikias and Raghuveer, 1987], provides information which conventional spectral estimation [Childers, 1978] fails to give. The applications of bispectral analysis are wide in range. Examples are found in oceanography, geophysics, passive sonar, biomedicine, telecommunications, speech processing, economic time series, fluid mechanics, plasma physics, and sunspot data analysis. Numerous references can be found in two extensive reviews by Lohmann and Wirnitzer [1984] and Nikias and Raghuveer [1987].

In the last 10 years or so, there has been major progress on the applications of bispectral analysis to astronomical and optical imaging [Bates, 1987; Lohmann and Wirnitzer, 1984; Weigelt, 1977; Bartelt *et al.*, 1984; Weigelt and Wirnitzer, 1983; Wirnitzer, 1985; Dainty and Northcott, 1986; Karbelkar and Nityananda, 1987]. Speckle imaging [Bates 1982], which stems directly from the invention of astronomical speckle interferometry by Labeyrie [1970] [cf. Labeyrie 1976] and the subsequent suggestion of speckle holography by Bates *et al.* [1973] and Liu and Lohmann [1973], has been significantly enhanced by bispectral processing [Weigelt and Baier, 1985].

The evaluation of the bispectrum with the aid of a digital computer, is, generally speaking, the numerical realization of the mathematical formulas. However, computational algorithms can be made more efficient by taking advantage of the various symmetries of the bispectrum [Lohmann and Wirnitzer, 1984]. Like ordinary correlation, either the triple correlation, or its Fourier transform (the bispectrum), can be realized by optical means. The majority of the proposed methods for realizing triple correlation via optical computing rely on acousto-optics [Lohmann and

Wirnitzer, 1984].

The usual reasons given for invoking the bispectrum in signal/image processing are its statistical properties. In power spectral estimation, the process under consideration is treated as a superposition of statistically uncorrelated harmonic components. The distribution of power among these frequency components is then estimated. As such, only average mechanisms governing the process are investigated since phase relationships between frequency components are suppressed [Huber *et al.*, 1971]. The information contained in the power spectrum is essentially that which is present in the autocorrelation [Bates and McDonnell, 1986]. This suffices for the complete statistical description of a Gaussian process of known mean [Haykin, 1983]. However, there are practical situations where one must look beyond the power spectrum (or autocorrelation) to obtain information regarding deviations from Gaussianness and the presence of nonlinearities. In fact, many real-world image formation procedures are far from Gaussian. By contrast, the bispectrum contains such information. The motivation for invoking the bispectrum in signal processing can thus be considered threefold: 1) to extract information due to deviations from Gaussianness, 2) to estimate the phase of non-Gaussian parametric signals, 3) to detect and characterize the nonlinear properties of instantaneous mechanisms which generate intermodulation products between their harmonic components [Nikias and Raghuveer, 1987].

Like the ordinary spectrum, the bispectrum can also be applied to the task of motion estimation since the bispectrum is shift-invariant. The major difference between the Fourier transform and the bispectrum is the following. When an object shifts and rotates, the bispectrum merely rotates, whereas the Fourier spectrum alters in a more complicated fashion. The magnitude of the Fourier spectrum does exhibit the same rotation as the object (see §2.2.3). However, the phase of the Fourier transform is affected by the translation as well as the rotation, implying that the complex Fourier spectrum is not suitable for shift-rotation motion analysis. It is shown in Chapter 5 that spectral phase plays an important role in rotation estimation. Since both the magnitude and the phase of bispectrum are shift invariant (as discussed below in §3.2.2), the phase can be employed for shift-rotation estimation, thereby improving the estimation performance by comparison with the standard Fourier intensity. This is the major motivation for invoking the bispectrum for shift-rotation motion estimation.

This chapter provides the theoretical background for the application of the bispectral technique to motion analysis. §3.2.1 introduces the bispectrum in the context of *higher order spectra.* §3.2.2 establishes the shift-rotation invariant properties of the bispectrum essential for rotational estimation. In §3.3, the technique, here called *angular bispectral correlation*, for estimating rotational motion is developed.

3.2 Theoretical Preliminaries

3.2.1 Bispectra and High Order Spectra

The widely applied conventional autocorrelation (see §2.2.1) is a special case of a sequence of autocorrelations. It is also termed the second order autocorrelation because the integrand in (2.11) is a product of a function and a shifted version of itself. A general n-th order autocorrelation of a 2-D real function can be defined by

$$f^{(n)}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{n-1}) = \int f(\mathbf{x}) f(\mathbf{x} + \mathbf{x}_1) f(\mathbf{x} + \mathbf{x}_2) ... f(\mathbf{x} + \mathbf{x}_{n-1}) dx dy \tag{3.1}$$

where $\mathbf{x}$ and $\mathbf{x}_i$ $(i = 1, 2, ..., n-1)$ are spatial position vectors (as defined in §1.3.1). The Fourier transform of the n-th order autocorrelation is generally called the n-th order spectrum. Spectra of order greater than 2 are termed higher order spectra, which are also known as polyspectra. Among the polyspectra, the third order spectrum, which is the Fourier transform of the third order autocorrelation or triple correlation, is conventionally termed the bispectrum. It has also been called the "Erlangen bispectrum" by Bates [1987].

Taking $n = 3$ in (3.1), the (auto) triple correlation $f^{(3)}$ is given by

$$f^{(3)}(\mathbf{x}_1, \mathbf{x}_2) = \int f(\mathbf{x}) f(\mathbf{x} + \mathbf{x}_1) f(\mathbf{x} + \mathbf{x}_2) dx dy \tag{3.2}$$

The bispectrum $F(\mathbf{u}_1, \mathbf{u}_2) \longleftrightarrow f^{(3)}(\mathbf{x}_1, \mathbf{x}_2)$ of a complex function $f(\mathbf{x})$ can be expressed [Lohmann and Wirnitzer, 1984] by

$$F(\mathbf{u}_1, \mathbf{u}_2) = F(\mathbf{u}_1) F(\mathbf{u}_2) F(-\mathbf{u}_1 - \mathbf{u}_2) \tag{3.3}$$

where $F(\mathbf{u}) \longleftrightarrow f(\mathbf{x})$.

The bispectrum has two inherent symmetries [Lohmann and Wirnitzer, 1984], i.e.

$$F(\mathbf{u}_1, \mathbf{u}_2) = F(\mathbf{u}_2, \mathbf{u}_1) = F(-\mathbf{u}_1 - \mathbf{u}_2, \mathbf{u}_1) \tag{3.4}$$

which follows directly from (3.3). When the image is real, the spectrum $F(\mathbf{u})$ is Hermitian, i.e.

$$F(-\mathbf{u}) = F^*(\mathbf{u}) \tag{3.5}$$

where the asterisk denotes complex conjugation. Therefore, when $f(\mathbf{x})$ is real, (3.3) can be rewritten as

$$F(\mathbf{u}_1, \mathbf{u}_2) = F(\mathbf{u}_1) F(\mathbf{u}_2) F^*(\mathbf{u}_1 + \mathbf{u}_2) \tag{3.6}$$

It is seen that the bispectrum is also Hermitian [Lohmann and Wirnitzer, 1984]. Since images considered in this thesis are real, (3.6) is understood from now on to define the bispectrum.

It is clear from (3.6) that the bispectrum is a four-dimensional (4-D) function. It is now convenient to rewrite $\mathbf{u}_1$ and $\mathbf{u}_2$, respectively, as $\mathbf{u}$ and $\boldsymbol{\alpha}$, which have respective Cartesian components (u, v) and (α, β), and respective polar coordinates $(\rho; \phi)$ and

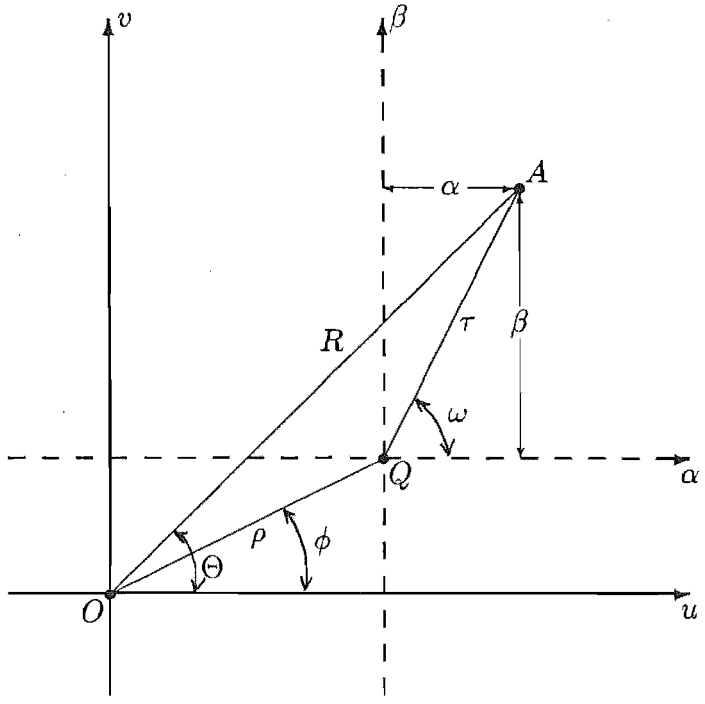


Figure 3.1. Polar and Cartesian coordinates of position vectors **u** and **α** in terms of which the bispectrum is conveniently defined.

$(\tau;\omega)$. Expressing the position vectors in terms of these particular coordinates, (3.6) can be written as

$$F(u,v,\alpha,\beta) = F(u,v)F(\alpha,\beta)F^*(u+\alpha, v+\beta) \tag{3.7}$$

for the Cartesian coordinate system, and

$$F(\rho;\phi;\tau;R) = F(\rho;\phi)F(\tau;\omega)F^*(R;\Theta) \tag{3.8}$$

for the polar coordinate system, where $(R;\Theta)$ are the polar coordinates of the point defined by the position vector $(u+\alpha, v+\beta)$, as is made clear in Fig. 3.1. The triangle OAQ, shown in Fig. 3.1, represents the geometric relation of **u** and **α** to **(u + α)**. Only four out of the total of six polar coordinates are needed to completely specify the triangle. In what follows, ρ, τ, R are invoked to specify the shape and the size, while ϕ specifies the orientation, of the triangle.

In the 4-D bispectral space, a position is specified by either four Cartesian coordinates (see (3.7)), or four polar coordinates (see (3.8)). It is also clear from Fig. 3.1 that these four coordinates represent a triangle and its orientation in the spectral domain. It is seen in the next subsection that when the object rotates, such a triangle remains congruent to itself, and only its orientation (ϕ) changes accordingly. This concept makes it possible to invoke bispectral functions for practical object motion estimation, as discussed later in §3.3.

3.2.2 Shift-Rotation Invariant Properties of the Bispectrum

In Chapter 2, in order to analyse motion, the shift-rotation invariant properties of the Fourier transform are studied. For the same reason, the shift-rotation invariant properties of the bispectrum [Bates, 1987] are discussed here.

Recalling what is discussed in §2.2.3, appealing to (2.13) and (3.7), and denoting a translated object by $g(x,y) = f(x',y')$, where x' and y' are the translationally transformed coordinates defined by (1.2), it is seen that the bispectrum of the translated object is given by

$$\begin{aligned} G(u,v,\alpha,\beta) &= F(u,v)\exp[i2\pi(ua+vb)] \\ &\quad F(\alpha,\beta)\exp[i2\pi(\alpha a+\beta b)] \\ &\quad F^*(u+\alpha,v+\beta)\exp[i2\pi((u+\alpha)a+(v+\beta)b)] \\ &= F(u,v)F(\alpha,\beta)F^*(u+\alpha,v+\beta) \\ &= F(u,v,\alpha,\beta) \end{aligned} \tag{3.9}$$

which confirms that the bispectrum is unchanged when the object translates.

Consider an object which only rotates. Referring to (2.18) and (3.8), and denoting the rotated object by $g(x,y) = f(x',y')$, where x' and y' are the rotationally transformed coordinates defined by (1.3), the bispectrum is then given by

$$G(\rho;\phi;\tau;R) = F(\rho;\phi+\psi)F(\tau;\omega+\psi)F^*(R;\Theta+\psi) = F(\rho;\phi+\psi;\tau;R) \tag{3.10}$$

which not only confirms that the bispectrum merely rotates through the same angle ψ as the object, but also indicates that the rotation can be characterized by the single coordinate ϕ. It is therefore understood that the triangle OAQ shown in Fig. 3.1 remains congruent to itself (i.e. the lengths of its sides do not change) as the object rotates (i.e. as ϕ varies). It is now convenient to define

$$F_{\rho,\tau,R}(\phi) = F(\rho;\phi)F(\tau;\omega)F^*(R;\Theta) \tag{3.11}$$

which emphasises the geometry of the triangle OAQ shown in Fig. 3.1. The subscripts imply that the lengths of the sides of triangle are fixed. Therefore, one can also employ the bispectrum values determined by any triplet ρ,τ,R to investigate the rotational behaviour of the bispectrum, e.g.

$$G_{\rho,\tau,R}(\phi) = F_{\rho,\tau,R}(\phi+\psi) \tag{3.12}$$

where $G_{\rho,\tau,R}$ is defined by replacing F by G in (3.11). The concept of the "congruent triangle" illustrates the rationale for introducing the bispectral angular correlation function discussed in §3.3.

Now consider an object which translates as well as rotates. Although the complex Fourier spectrum of the object depends on the order of the two operations (see §2.2.3), the bispectrum does not. This can be seen by substituting (2.20) and (2.21) respectively, into (3.7), i.e.

$$\begin{aligned} G(u,v,\alpha,\beta) &= F(u',v')\exp[i2\pi(u\hat{a}+v\hat{b})] \\ &\quad F(\alpha',\beta')\exp[i2\pi(\alpha\hat{a}+\beta\hat{b})] \\ &\quad F^*(u'+\alpha',v'+\beta')\exp[i2\pi((u+\alpha)\hat{a}+(v+\beta)\hat{b})] \\ &= F(u',v')F(\alpha',\beta')F^*(u'+\alpha',v'+\beta') \\ &= F(u',v',\alpha',\beta') \end{aligned} \tag{3.13}$$

where, referring to §2.2.3, for rotation followed by translation,

$$\begin{aligned} \hat{a} &= a \\ \hat{b} &= b \end{aligned} \tag{3.14}$$

and for translation followed by rotation,

$$\begin{aligned} \hat{a} &= a' \\ \hat{b} &= b' \end{aligned} \tag{3.15}$$

and remembering that u', v', defined by (2.16), represent the rotated Cartesian coordinate system in Fourier space. Therefore, the expression for $G(u,v,\alpha,\beta)$ in an arbitrarily transformed coordinate system depends solely upon primed arguments, which confirms the claim made in the fifth paragraph of §3.1 that the bispectrum is an ideal rotation estimator because it is shift-invariant. This is more obvious when (3.13) is expressed in polar coordinates since, by replacing the primed Cartesian coordinates with polar coordinates, the equation is of exactly the same form as (3.10). Consequently, (3.12) remains true for this situation.

In summary (compare with §2.2.3), when an object translates, both the magnitude and phase of the bispectrum are unaltered. When the object rotates, the bispectrum rotates by the same angle. When the body both rotates and translates, the bispectrum merely rotates. This rotational property of the bispectrum is equivalently illustrated by the congruent "rotating triangle" introduced earlier in this subsection.

3.3 Rotational Estimation via Bispectral Correlation

In §2.3.6, the angular correlation employing the Fourier transform magnitude is defined. The bispectral angular correlation is defined by substituting $|F(\rho;\phi)|$ and $|G(\rho;\phi)|$ with $F(\rho;\phi;\tau;R)$ and $G(\rho;\phi;\tau;R)$, respectively, in (2.44). However, calculating the bispectrum is in general computationally expensive, because the bispectrum is a 4-D complex function, even when the various symmetries (see §3.2.1) are noted. A restricted angular bispectral correlation can be defined by employing only enough information to unequivocally determine rotation. Referring to (3.12) and Fig. 3.1, the angular correlation is conveniently expressed in terms of $F_{\rho,\tau,R}(\phi)$ and

$G_{\rho,\tau,R}(\phi)$, which emphasise the congruent triangle concept introduced in §3.2.2, i.e.

$$\Gamma(F_{\rho,\tau,R}, G_{\rho,\tau,R}, \phi) = \left| \int_0^{2\pi} F_{\rho,\tau,R}(\phi')\, G^*_{\rho,\tau,R}(\phi' + \phi)\, d\phi' \right| \tag{3.16}$$

It is seen that $\Gamma(F_{\rho,\tau,R}, G_{\rho,\tau,R}, \phi)$ is effectively a 1-D correlation operation, since two 1-D complex functions are correlated with the variable ϕ which varies from 0 to 2π. In order to determine the rotation angle, the angular correlation domain is searched for an angle ψ such that $\Gamma(F_{\rho,\tau,R}, G_{\rho,\tau,R}, \psi)$ is a maximum. In the absence of noise on the data, ψ is the correct rotation angle.

A computational example is now presented to illustrate the principle of angular bispectral correlation. Fig. 3.2(a) shows an image of the letter **R**. Its shifted (upwards and rightwards) and rotated (by 45°) version is shown in Fig. 3.2(b). Fig. 3.2(c) shows the magnitude of the spectrum for the object shown in Fig. 3.2(a), with the triangle for the bispectrum shown highlighted. For both the "**R**" and its shifted and rotated version, the magnitudes of the complex bispectral functions for the triangle are plotted as functions of ϕ in, respectively, Figs. 3.3(a) and (b). The result of correlating these two 1-D functions is plotted in Fig. 3.3(c), and the correlation peak is seen to be situated at $\phi = \pi/4$.

To conclude this chapter, it is noted that practical considerations involving the implementation of angular correlation in the presence of noise are discussed in Chapter 5. In addition, a comparative study of the performances of the angular correlation methods employing the spectral and the bispectral angular correlations are presented in that chapter.

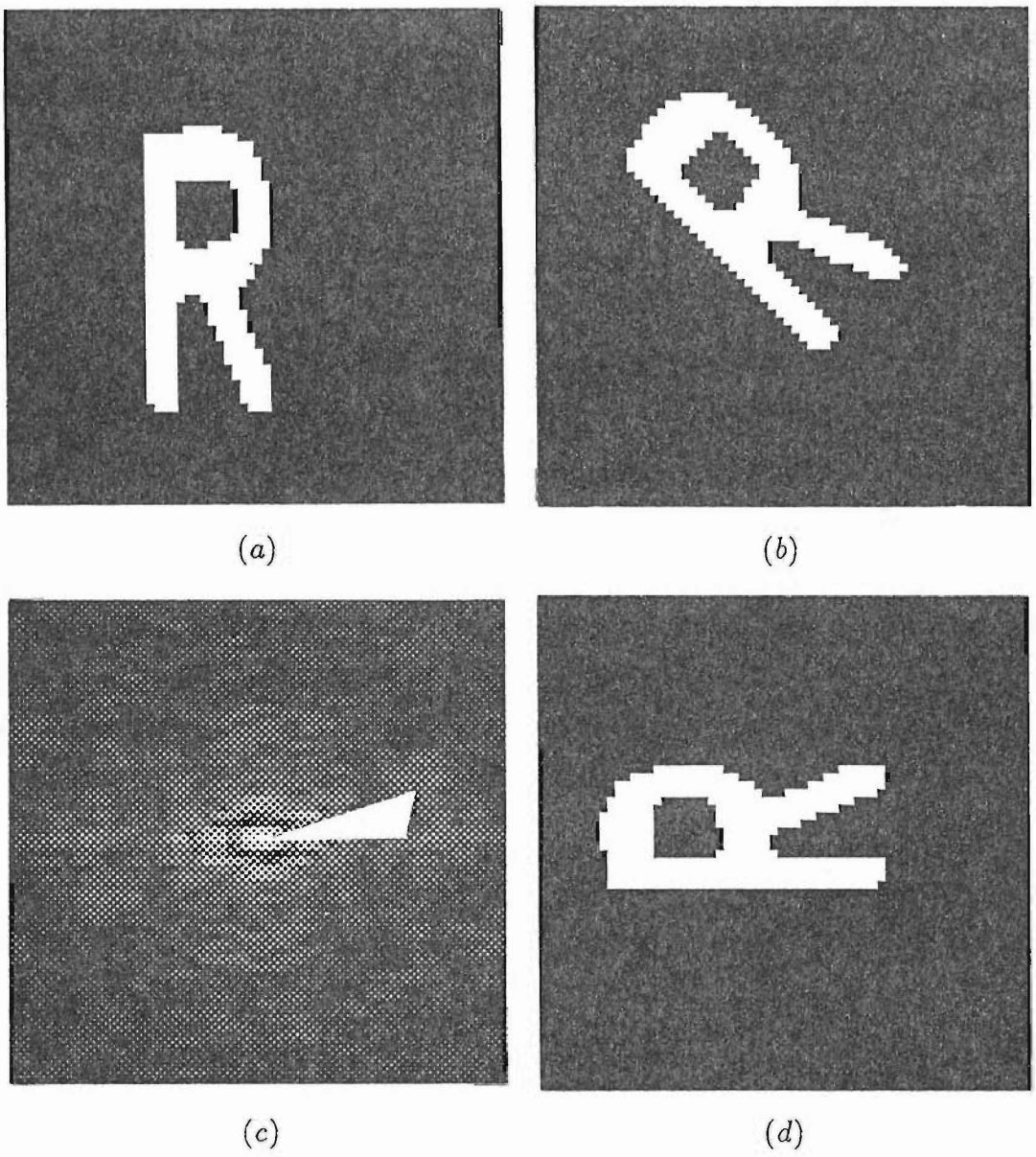


Figure 3.2. Illustration (I) of the principle of angular bispectral correlation: (a) the original object "R"; (b) the shifted and rotated version of the object shown in (a); (c) the magnitude of the spectrum of the object shown in (a), and the triangle for the bispectrum is highlighted; (d) the shifted (8 pixels upward and 8 pixels rightward) and then rotated ($\pi/2$) version of (a). Note that (a), (c) and (d) are invoked in Chapter 5.

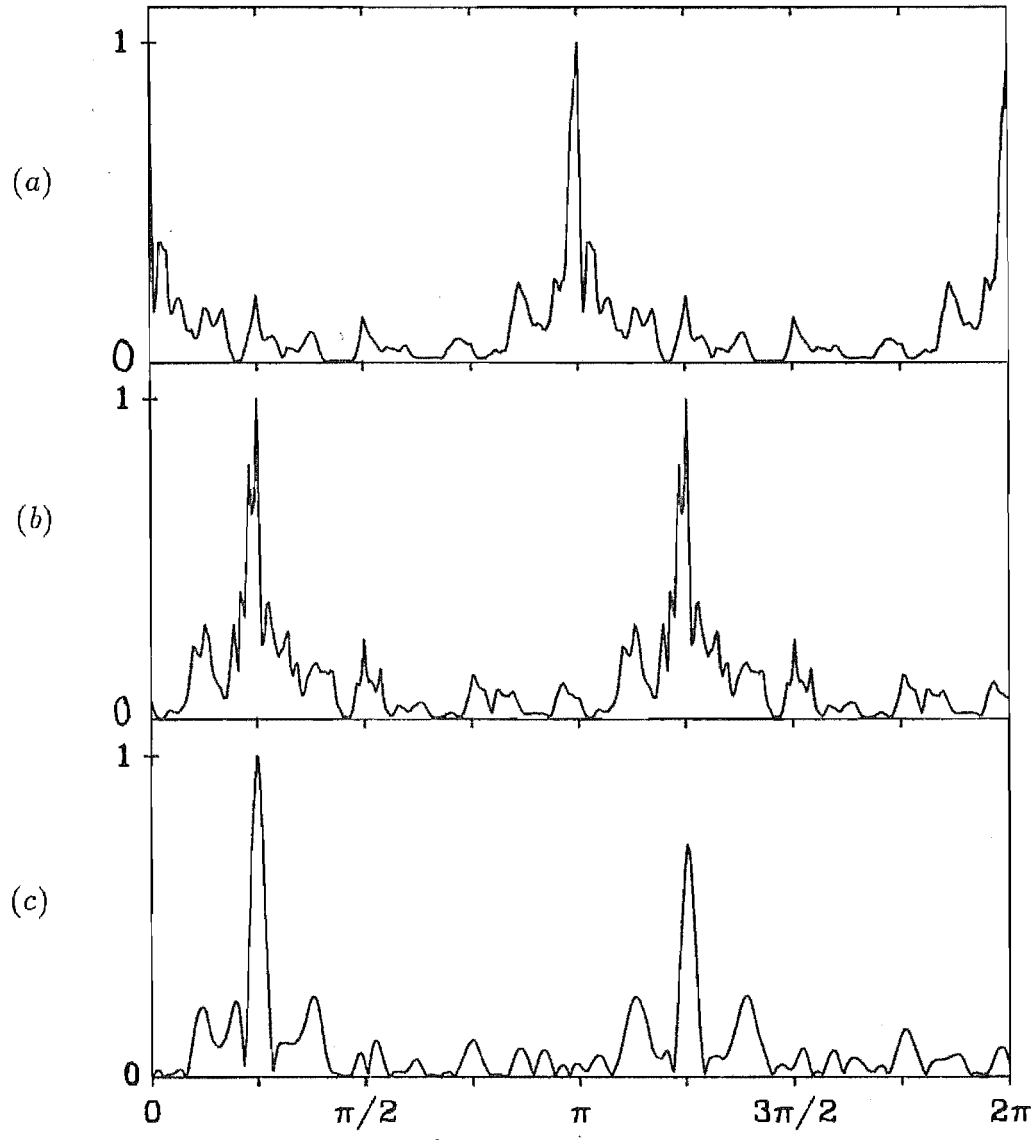


Figure 3.3. Illustration (II) of the principle of angular bispectral correlation: (a) the magnitude of the complex bispectral function as a function of ϕ for the "R" shown in Fig. 3.2(a); (b) the magnitude of the complex bispectral function as a function of ϕ for the "R" shown in Fig. 3.2(b); (c) the result of correlating (a) and (b). The correlation peak in (c) is situated at $\phi = \pi/4$. Note that all plotted curves represent normalized functions, i.e. the maximum of each of them is set to 1.

Chapter 4

Implementation of Phase Correlation for Processing Video Image Sequences

This chapter discusses three aspects of processing video image sequences by the phase correlation technique introduced in §2.3.4. Firstly, the effect of video camera lag [Batchelor *et al.*, 1985; Garrett and Smithson, 1987] (see also §1.2.1 and Chapter 6) and its compensation are described in §4.1. Secondly, problems which have to be solved in order to implement the phase correlation technique for processing video image sequences are discussed in §4.2. Thirdly, three modifications to the phase correlation method needed to compensate for various defects of real world data are presented in §4.3. Algorithms for processing image sequences are presented in §4.4, with several computational results. Since the processing of video image sequences is the main concern of this chapter, the terms "frame" and "image" are used interchangably to refer to a single image from a sequence.

4.1 Compensating for Camera Lag

It is now standard for many machine vision systems to obtain images from video sources. For example, an industrial inspection system or a robot vision system commonly employs a TV camera as its sensor. Medical examination conventionally employs X-ray radiographic systems which now combine the use of image intensifiers and video cameras as image acquiring devices. What are given particular attention in this thesis are the coronary angiographic (CA) images obtained from such an X-ray imaging system, often referred to specifically as an X-ray angiography system (see Chapter 7 for details of how coronary angiographic image sequences are obtained). The effects of video camera lag on the image of a rotating and translating object are examined in §4.1.1. It is demonstrated in §4.1.2 that, in the absence of all recording imperfections, weighted subtraction of successive TV images can generate a true image of an object which both translates and rotates. Finally, in §4.1.3, three problems, which have to be considered if one wishes to process CA images optimally, are

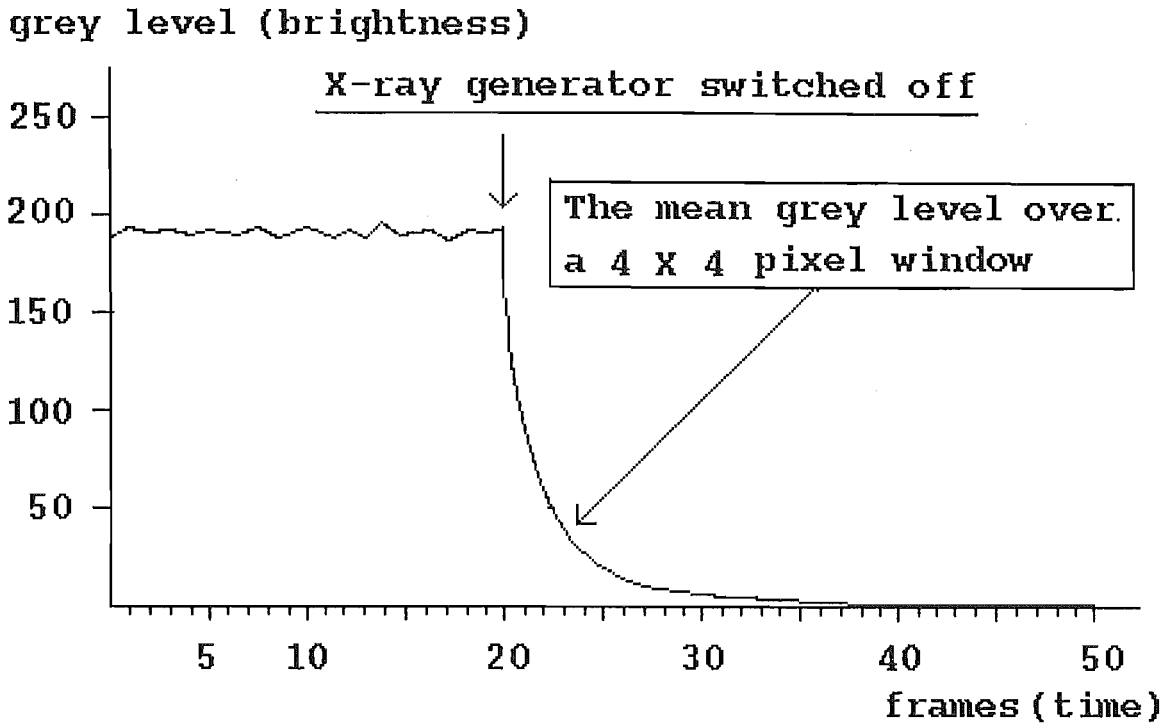


Figure 4.1. The curve of the *brightness-vs-time* of a vidicon camera, before and after the X-ray tube has been turned off. It shows that the brightness decays exponentially. The brightness at a particular time is obtained by averaging the pixel values from a 4×4 pixel region of the video image. Note that the interval between two consecutive frames is $40msec$.

SEE ERRATA disscussed.

4.1.1 Effect of Video Camera Lag

The video cameras employed in most X-ray angiographic systems are of the vidicon or plumbicon types, both of which exhibit appreciable "lag", i.e. once excited, a spot on the camera target screen persists for tens or hundreds of milliseconds [Batchelor *et al.*, 1985; Garrett and Smithson, 1987]. It can be shown that the brightness of the video screen decays exponentially by plotting the *brightness-vs-time* curve (see Fig. 4.1) during an interval of time within which the X-ray tube is turned off.
SEE ERRATA Table 4.1.1 lists the lag constants for several commercial camera systems commonly in use.

The image-intensifier is excited each time the X-ray tube is pulsed. Consequently, the image appearing on the camera target screen is effectively a multiple exposure. When the object moves during recording, the resultant images are noticeably blurred, as is illustrated in Fig. 4.2 and Fig. 4.3 (Note that the use of a coronary phantom,

Camera type	Vidicon (normal)	Vidicon (lead oxide)	Vidicon (multilayer)
Persistence after 200ms	15 %	2.0 %	8.0 %

Table 4.1. Some typical vidicon camera lags.

the images of which are shown in Fig. 4.2, is discussed in Appendix A).

Figs. 4.2(a) and (b) show two sequential images of a contrast-medium-filled coronary phantom recorded while the phantom is moving. Fig. 4.2(c) shows an image of the same phantom when it is stationary. Note that camera lag has caused a significant blurring of the edges of the phantom in Figs. 4.2(a) and (b) as compared to Fig. 4.2(c). Figs. 4.3(a), (b) and (c) are three frames in an actual CA image sequence. In Fig. 4.3, while (a) and (b) are two successive frames recorded when the artery segment is in a state of fast motion, (c) is the one which is recorded when the artery is nearly stationary (corresponding to the end of the ventricular filling period).

Consider a sequence of "ideal" TV images, uncontaminated by recording noise or any other distortion. $k_{m,n}$ is taken to denote the fraction of intensity, due to the $(m-n)^{th}$ X-ray pulse, persisting on the target screen at the instant of the m^{th} pulse; implying that $k_{m,0} = 1$ for each m. Also consider an object which translates and rotates while the image sequence is recorded, with $a(\mathbf{x})$ being the true image of the object (i.e. what would always be recorded if the object was stationary and if the target screen exhibited no lag), where $\mathbf{x}$ is the position vector of an arbitrary point in image space. Denoting the translation and rotation of the object (with respect to any fixed axes), at the instant of the m^{th} X-ray pulse, by the shift $\mathbf{s}_m = (a_m, b_m)$ and the dyadic $\mathbf{R}_m$ (see §1.3.1) respectively, the m^{th} CA image $f_m(\mathbf{x})$ can be expressed as

$$f_m(\mathbf{x}) = \sum_{n=0}^{m} k_{m,n} a(\mathbf{R}_{m-n} \cdot (\mathbf{x} + \mathbf{s}_{m-n})) \tag{4.1}$$

which is, effectively, a multiple exposure. In practice, X-ray tubes are regularly pulsed. The video frames are also recorded periodically. Here, the X-ray pulse repetition frequency is taken to be equal to the video frame rate, implying that,

$$k_{m,n} = k^n \tag{4.2}$$

where $k = k_{m,1}$ is a constant for a given video camera. For example, consider the calculation of k for the normal Vidicon camera (refer to Table 4.1.1). Since it usually takes 40 millisecond (ms) to scan the video signal for a full frame, 200ms is the time taken for scanning 5 frames. Therefore, $k^5 = 0.15$, which gives $k = 0.684$.

In practice, k may be estimated by observing the response of the camera to a step change in X-ray intensity. It is found that the simplest method is to record a sequence of images of uniform phantom material and record the video signal immediately following switching off the X-ray generator. The mean intensity of a region at the

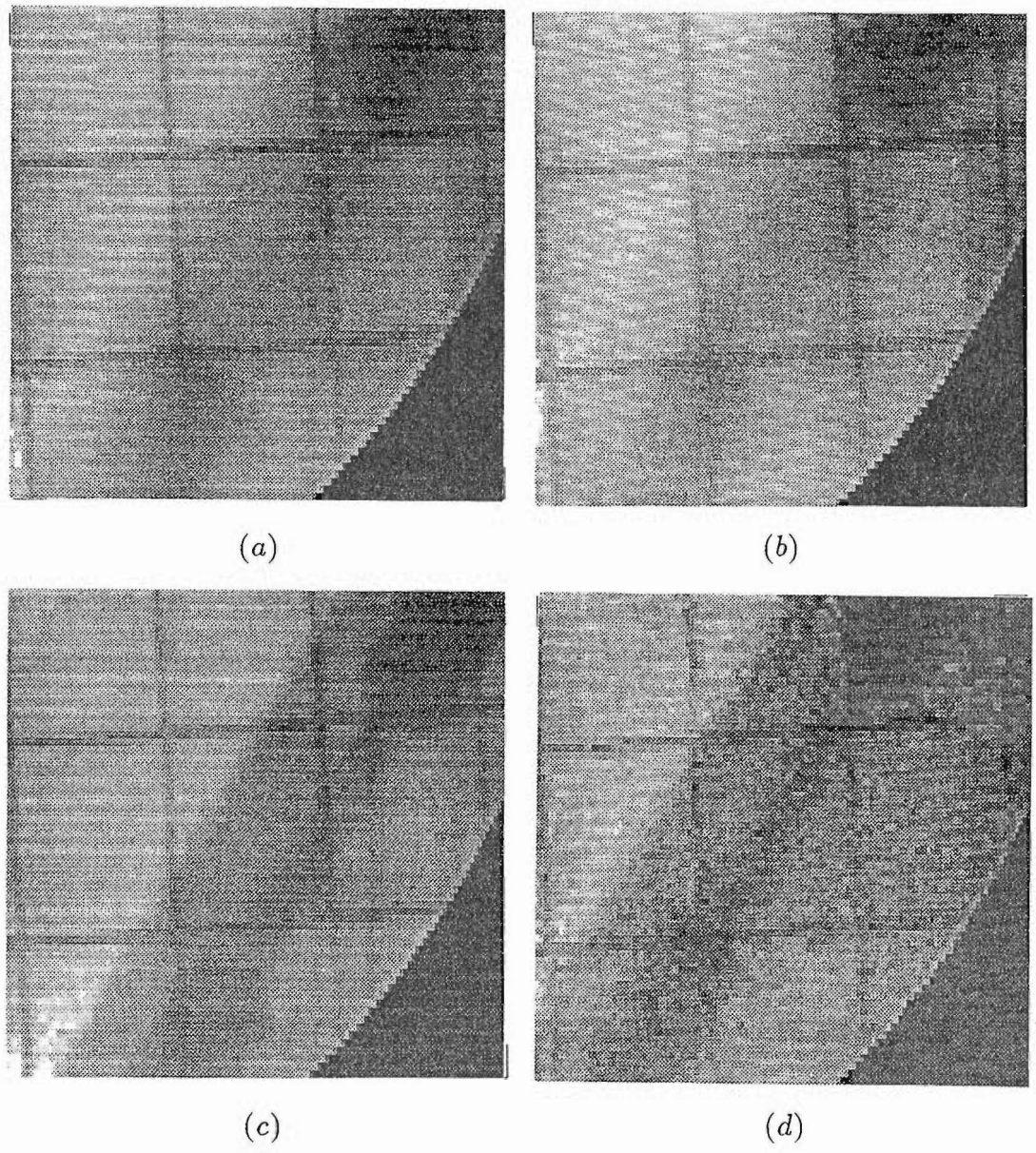


Figure 4.2. Four 128 × 128 pixel (X-ray) images of a contrast-medium-filled coronary artery phantom illustrating the blurring effect of the lag on a moving object: (a) and (b) consecutive frames recorded during motion of the phantom towards the top left corner; (c) a frame recorded when the phantom is stationary; (d) result of the subtraction of (a) from (b) for $k = 0.7$ (see (4.3)).

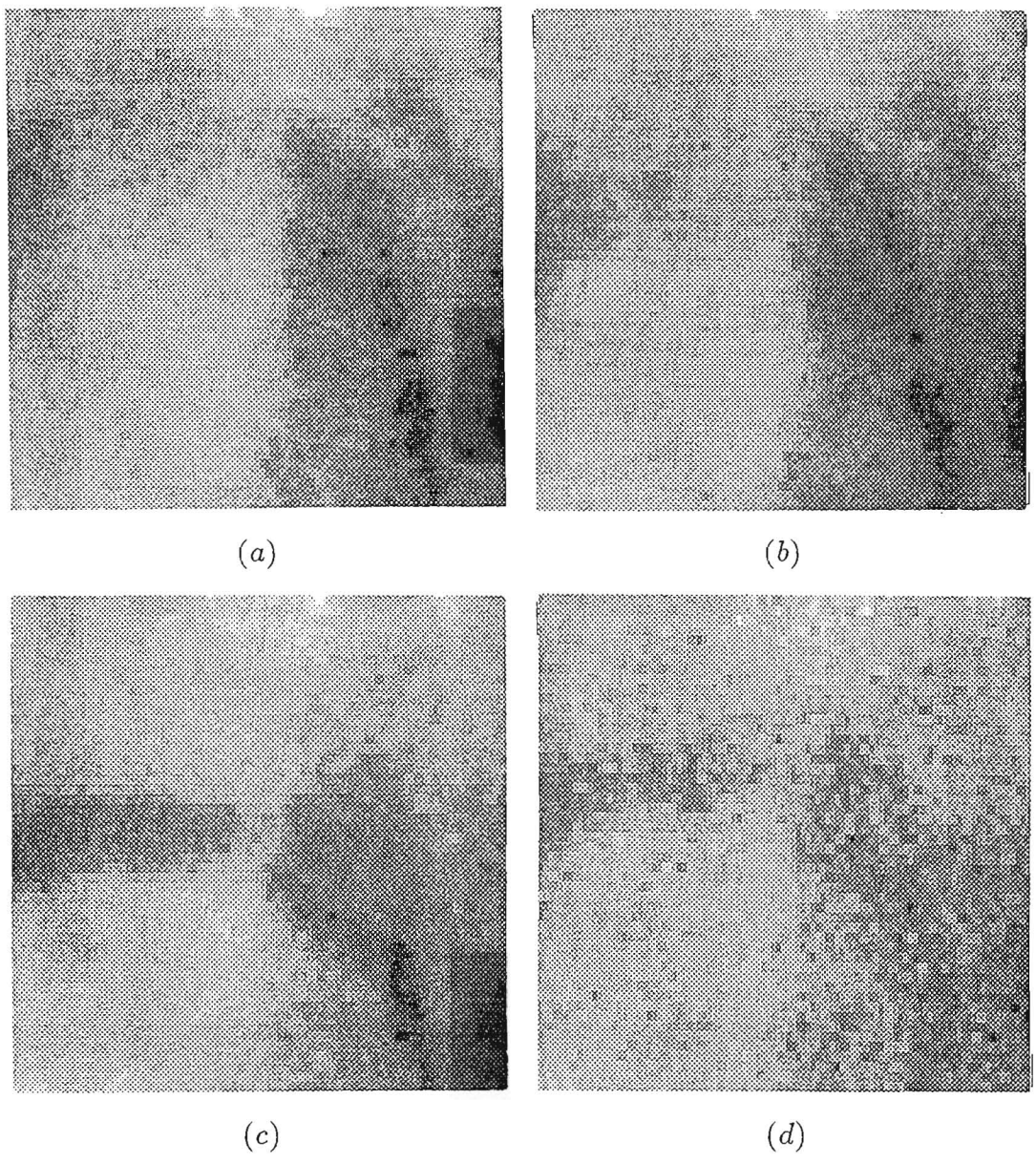


Figure 4.3. Illustration of the blurring effect of the lag on a moving contrast-medium-filled coronary artery segment: (a) frame 3 of a sequence of coronary angiograms; (b) frame 4 from the same sequence; (c) frame 9 from the sequence, recorded when the artery is stationary; (d) result of the subtraction of (a) from (b) for $k = 0.6$ (cf (4.3)).

SEE ERRATA

centre of each digitised frame is then plotted as a function of time (e.g. as shown in Fig. 4.1) and k is estimated by curve fitting.

In clinical practice, the repetition frequency of X-ray pulses is equal to the cine film frame rate but not necessarily to the video frame rate. This reflects the importance placed on the cine film as the main recording medium in a conventional angiography system (see Chapter 6). For example, with the CGR cineangiographic system currently used at the Cardiology Department, Princess Margaret Hospital, Christchurch (see Chapter 7), available cine frame rates range from 20 to $75 fps$ (*frames per second*). One of the available rates is $50 fps$, which coincides with the interlaced video field rate of 50 fields/sec while several other available rates are close to this rate, for example $45 fps$. Kruger [1986] has described the synchronization between X-ray pulsation and video frame scanning rate in recently developed digital angiographic systems. However, experience with recording shows that asynchronous pulse rates can be tolerated to a certain extent. This is demonstrated in Figs. 4.2 and 4.3 which correspond to X-ray pulse repetition and video frame rates that are not equal to each other. The results (shown in Figs. 4.2(d) and 4.3(d)) for the compensation scheme presented in the next sub-section indicate that X-ray pulse repetition and video frame rates of standard video apparatus can be considered, in practice, to be effectively equal.

4.1.2 Subtraction of Successive TV Images

It can be seen from (4.1) and (4.2) that

$$f_{m+1}(\mathbf{x}) - k f_m(\mathbf{x}) = a(\mathbf{R}_{m+1} \cdot (\mathbf{x} + \mathbf{s}_{m+1})) \tag{4.3}$$

which confirms that, in principle, subtraction of successive TV images produces a true image of the object. This image is translated (i.e. shifted) by $\mathbf{s}_{m+1}$ and is rotated by the angle implicit in $\mathbf{R}_{m+1}$. Fig. 4.2(d) illustrates the result of applying (4.3) to the coronary phantom images shown in Figs. 4.2(a) and 4.2(b). Similarly, Fig. 4.3(d) shows the result of the compensation scheme on the actual CA images shown in Figs. 4.3(a) and (b). It is evident that the resolution of the images resulting from the subtraction is comparable to that of the images of static objects, as confirmed by comparisons between Figs. 4.2(c) and 4.2(d), and between Figs. 4.3(c) and 4.3(d). While this subtraction operation often reveals detail that is barely apparent in unprocessed CA images, the version of the true image so generated tends to be distinctly noisy (see Fig. 4.2(d)) [Balter *et al.*, 1984]. Due to the random nature of the noise in the CA images (see Chapter 6), the result of the subtraction procedure on LHS of (4.3) has a noise variance which is higher than in any typical CA image. Assuming the noise is Gaussian (see §6.2.4), the variance of the noise in the subtracted image is, from (4.3),

$$\sigma^2 = (1 + k)\sigma_f^2 \tag{4.4}$$

where σ_f^2 is the noise variance of a typical CA image in a sequence. It is reasonable to assume that all CA images captured under the same clinical conditions have the same noise statistics.

Although it is desirable to remove the effect of lag on images of moving objects before applying any image processing techniques, the lag effect can be advantageous in certain clinical applications [Davison and *et. al*, 1977]. First of all, it improves the signal-to-noise of X-ray images because (4.1) represents a weighted average. Furthermore, because human eyes are more sensitive to moving objects than to static ones [Cornsweet, 1971], when video images are viewed, the lag's blurring effect is usually not noticable for objects that are not moving quickly. It is only when a video image is arrested that the blur is seen. This has been the experience of clinicians and researchers in the Cardiology Department, Princess Margaret Hospital, Chirstchurch. An implication arising from the above is that a TV camera with significant lag is preferred to one with less lag for monitoring an operation.

The problems which must be solved to produce more faithful images are outlined in the next sub-section.

4.1.3 Exposure, Enhancement, and Registration

It is difficult to expose different CA images identically for many reasons, perhaps the chief being the variable intensity of the X-ray pulses. This implies that the m^{th} CA image is not $f_m(\mathbf{x})$, as defined by (4.1), but $(e_m f_m(\mathbf{x}))$, where the m^{th} exposure factor e_m varies in general with m. The solution to this exposure problem is to multiply the m^{th} CA image by a constant before performing the subtraction operation described in §4.1.2 and, then, to vary this constant until the outcome of the subtraction is "most like" the image of a single object. Because such compensation for variable exposure can never be perfect, the residual unwanted detail in the subtracted images constitutes a species of noise, which adds to all the other forms of contamination that manifest themselves in practice.

Given a sequence of CA images, the compensation for multiple exposure blurring by the weighted subtraction generates a sequence of noisy versions of the true images of the object. To obtain an improved image one must contrive means of enhancing these noisy versions. The obvious solution to this enhancement problem is to appropriately superimpose all of the noisy versions.

Superimposing the noisy versions is only effective if the detail representing the object in each version is registered such that it is reinforced when the versions are superimposed. To solve this registration problem one needs to deduce the rotations and shifts experienced by the object between the instants at which the X-ray tube is pulsed. The next section presents a solution for an object whose motion is purely translational.

4.2 Phase Correlation

The principle of the phase correlation technique has been introduced in Chapter 2. However, the form of phase correlation presented in §2.3.4 can only be invoked for processing two images between which the object has shifted only a relatively small

distance. It is not appropriate for automatic processing of most practical video image sequences where the object's shift between any two images can be so large that the estimation of shift is prone to the kinds of error discussed in §2.3.4. The new form of phase correlation introduced in this section incorporates a predictive estimation technique which is suitable for processing image sequences. This technique is, from now on, termed the *sequential predictive phase correlation* (SPPC). Consequently, in the rest of this thesis, the PC technique only refers to the step in SPPC for estimating the shift of an object between a single pair of images in a sequence.

The mathematical formulation of SPPC is presented in §4.2.1. Associated with the implementation of the technique are two critical problems, i.e. how sensitive the phase correlation is to, respectively, multiple exposure distortion and rotation of the translating object. These two problems are discussed in §4.2.2.

4.2.1 Processing Image Sequences

For a purely translating object (i.e. the motion is such that each $\mathbf{R}_m$ is the unit dyadic), the outcome of the subtraction operation (4.3) is, in the absence of any of the imperfections which are manifested in practice,

$$a_{m+1}(\mathbf{x}) = a(\mathbf{x} - \mathbf{s}_{m+1}) \tag{4.5}$$

which is the $(m+1)^{th}$ translated version of the object's true image.

Defining $a(\mathbf{x}) \longleftrightarrow A(\mathbf{u})$ and $a_m(\mathbf{x}) \longleftrightarrow A_m(\mathbf{u})$, and recalling (4.5) and the shift theorem for Fourier transforms (see §2.2.1), it is seen that

$$A_m(\mathbf{u}) = A(\mathbf{u}) \exp(i2\pi\mathbf{u} \cdot \mathbf{s}_m) \tag{4.6}$$

It follows from (4.6) that

$$\Phi_m(\mathbf{u}) = \text{Phase}\{A_m(\mathbf{u})\} - \text{Phase}\{A(\mathbf{u})\} = 2\pi\mathbf{u} \cdot \mathbf{s}_m \tag{4.7}$$

and

$$\Phi_{m+1,m}(\mathbf{u}) = \text{Phase}\{A_{m+1}(\mathbf{u})\} - \text{Phase}\{A_m(\mathbf{u})\} = 2\pi\mathbf{u} \cdot [\mathbf{s}_{m+1} - \mathbf{s}_m] \tag{4.8}$$

Equation (4.7) can be invoked for estimating the shifts $\mathbf{s}_m$ of an object in an image sequence with respect to the object's position in $a(\mathbf{x})$. For convenience, the shift so estimated is here termed the absolute shift (AS). Alternatively, (4.8) can be used for estimating the shift of the object between two successive frames of a sequence, termed the relative shift (RS).

Ideally, inverse Fourier transformation of the phase difference spectra, i.e. $\exp(i\Phi_m(\mathbf{u}))$ and $\exp(i\Phi_{m+1,m}(\mathbf{u}))$, yield, respectively, $\delta(\mathbf{x}-\mathbf{s}_m)$ and $\delta(\mathbf{x}-\mathbf{s}_{m+1}+\mathbf{s}_m)$, implying (naively) that the AS, $\mathbf{s}_m$, or the RS, $\mathbf{s}_{m+1,m} = (\mathbf{s}_{m+1} - \mathbf{s}_m)$, can be recovered with arbitrary precision. In practice, of course, an ideal δ-function cannot be achieved because of various imperfections of the real world data, in particular, the presence of noise, and also because of the effects of truncation. These considerations are examined in §4.3.

Two computational schemes for SPPC can be devised based on estimating the AS and the RS respectively. Before establishing these two schemes, it is instructive to discuss characteristics of the motion of coronary arteries (cf. Chapter 6). Since cardiac motion is quasi-periodic, the maximum movement of any part of the coronary arteries can be approximately observed within a single cycle, although artery segments on different areas of the heart surface have different amounts of motion. The movement of a certain artery segment (feature of clinical interest (FOI)) which appears on a digitised image sequence could be as large as several tens of pixels. However, in practical computations, a rectangular area of interest (AOI) of 64×64 pixels containing the FOI is usually chosen for reasons made clear later in §4.3. If one simply extracts data from a fixed AOI throughout the sequence, a large shift of the FOI between any two frames causes significant differences in the AOI content between these two frames. Subsequently, errors occur in estimating the shift, for the reasons given in §2.3.4. Therefore, in both of the schemes presented below, a predictive method is invoked for repositioning the AOI in each frame of the sequence.

The procedure for estimating the AS is now described. First, one particular image is chosen to be a reference from an image sequence $\{ a_m(\mathbf{x}), m = 1, 2, \cdots, M \}$. In general, any image in the sequence can be chosen as the reference, but for convenience of exposition, $a_1(\mathbf{x})$ is chosen here. In the following, the position of the AOI, denoted as $\mathbf{P}_m$, is defined to be the position of the bottom left corner of the AOI in $a_m(\mathbf{x})$. For the reference image $a_1(\mathbf{x})$, $\mathbf{P}_1$ is chosen so that the FOI lies centrally within the AOI. For other images in the sequence, assuming that $\mathbf{s}_m$ has already been estimated, the location of the AOI in $a_{m+1}(\mathbf{x})$ is

$$\mathbf{P}_{m+1} = \mathbf{P}_1 + \mathbf{s}_m \tag{4.9}$$

The AS, $\mathbf{s}_{m+1}$, is then computed by

$$\mathbf{s}_{m+1} = \hat{\mathbf{s}}_{m+1} + (\mathbf{P}_{m+1} - \mathbf{P}_1) \tag{4.10}$$

where $\hat{\mathbf{s}}_{m+1}$ is the shift vector directly estimated by applying phase correlation to the AOIs of $a_1(\mathbf{x})$ and $a_{m+1}(\mathbf{x})$.

The procedure for estimating RS differs from AS only in that shifts are estimated between consecutive images in a sequence. The method for extracting the AOI, for the RS scheme, is derived from (4.9) and (4.10). Knowing that $\mathbf{s}_{1,1} = \mathbf{s}_1 = (0, 0)$ (i.e. the "shift" between $a_1(\mathbf{x})$ and $a_1(\mathbf{x})$ is zero) and $\mathbf{s}_2 = \mathbf{s}_{2,1}$ (i.e. the AS between the reference $a_1(\mathbf{x})$ and $a_2(\mathbf{x})$ is the same as the RS between them), the AS $\mathbf{s}_m$ is seen to be

$$\begin{aligned} \mathbf{s}_m &= \mathbf{s}_2 + (\mathbf{s}_3 - \mathbf{s}_2) + (\mathbf{s}_4 - \mathbf{s}_3) + \cdots + (\mathbf{s}_m - \mathbf{s}_{m-1}) \\ &= \mathbf{s}_{2,1} + \mathbf{s}_{3,2} + \mathbf{s}_{4,3} + \cdots + \mathbf{s}_{m,m-1} \end{aligned} \tag{4.11}$$

which is the sum of all previously estimated relative shifts. From (4.9), $\mathbf{P}_m = \mathbf{P}_1 + \mathbf{s}_{m-1}$, so that

$$\begin{aligned} \mathbf{P}_{m+1} &= \mathbf{P}_1 + \mathbf{s}_m \\ &= (\mathbf{P}_m - \mathbf{s}_{m-1}) + \mathbf{s}_m \\ &= \mathbf{P}_m + (\mathbf{s}_m - \mathbf{s}_{m-1}) \\ &= \mathbf{P}_m + \mathbf{s}_{m,m-1} \end{aligned} \tag{4.12}$$

implying that the AOI position in $a_{m+1}(\mathbf{x})$ is that in $a_m(\mathbf{x})$ plus the RS $\mathbf{s}_{m,m-1}$. Similarly, referring to (4.10), (4.11) and (4.12), the RS $\mathbf{s}_{m+1,m}$ is given by,

$$\mathbf{s}_{m+1,m} = \hat{\mathbf{s}}_{m+1,m} + (\mathbf{P}_{m+1} - \mathbf{P}_m) \tag{4.13}$$

where $\hat{\mathbf{s}}_{m+1,m}$ is estimated from the AOIs in $a_m(\mathbf{x})$ and $a_{m+1}(\mathbf{x})$.

Experience shows that the RS scheme is superior to the AS scheme for estimating the motion of coronary arteries in an X-ray angiogram sequence. With the AS scheme, the shift estimates are usually sufficiently accurate for images which are captured shortly after the capturing of the reference image. However, the phase correlation technique applied to two images which are captured at significantly different times tends to a give poor result. One reason is that the artery's shape and orientation change relatively slowly. However, these changes accumulate through time. The shape change and rotation (two kinds of deformation from which CA image sequences suffer) are, therefore, less likely to affect the performance of the phase correlation technique applied to two consecutive images in a sequence than for the AS scheme. The next sub-section discusses the sensitivity of phase correlation to multiple exposure and rotation.

4.2.2 Sensitivity of Phase Correlation To Multiple Exposure and Rotation

It is shown in §4.2.1 that the SPPC technique can be invoked to infer an object's arbitrary translational motion in an image sequence which has been compensated for camera lag. However, it is argued in this sub-section that it may be advantageous, when the captured images are effectively multiple exposures, to perform phase correlation before camera lag compensation. When the exposures are multiple and the object translates uniformly, phase correlation is fully effective. Furthermore, my computational experience indicates that the phase correlation technique performs usefully for arbitrary translational motion, even together with some slight rotation of the object.

On defining $f_m(\mathbf{x}) \longleftrightarrow F_m(\mathbf{u})$, and recognising that $g(\mathbf{R}\cdot\mathbf{x}) \longleftrightarrow G(\mathbf{R}\cdot\mathbf{u})$ for arbitrary $g(\mathbf{x}) \longleftrightarrow G(\mathbf{u})$ and any rotation dyadic $\mathbf{R}$, it is seen from (4.1) and (4.2) that

$$F_m(\mathbf{u}) = \sum_{n=0}^{m} k^n A(\mathbf{R}_{m-n}\cdot\mathbf{u}) \exp(i2\pi \mathbf{R}_{m-n}\cdot\mathbf{u}\cdot\mathbf{s}_{m-n}) \tag{4.14}$$

so that

$$\begin{aligned} F_{m+1}(\mathbf{u}) &= F_m(\mathbf{u}) \exp(i2\pi\mathbf{u}\cdot[\mathbf{s}_{m+1} - \mathbf{s}_m]) \\ &+ \sum_{n=0}^{m} k^n \left[A(\mathbf{R}_{m+1-n}\cdot\mathbf{u}) \exp(i2\pi\mathbf{R}_{m+1-n}\cdot\mathbf{u}\cdot\mathbf{s}_{m+1-n})\right. \\ &\quad \left. - A(\mathbf{R}_{m-n}\cdot\mathbf{u}) \exp(i2\pi[\mathbf{R}_{m-n}\cdot\mathbf{u}\cdot\mathbf{s}_{m-n} + \mathbf{u}\cdot\mathbf{s}_{m+1} - \mathbf{u}\cdot\mathbf{s}_m])\right] \\ &+ k^{m+1} A(\mathbf{R}_0\cdot\mathbf{u}). \end{aligned} \tag{4.15}$$

It seems that there is no benefit to be expected in general from attempting to interpret the difference of the phases of $F_{m+1}(\mathbf{u})$ and $F_m(\mathbf{u})$ in the same way that $\Phi_{m+1,m}(\mathbf{u})$ is

interpreted in §4.2.1. On the other hand, when k is small enough, and the differences between $\mathbf{R}_{n+1}$ and $\mathbf{R}_n$ are sufficiently slight (for any n), one sees that the first term on the right hand side of (4.15) may dominate the final term (the latter tends to be negligible when m exceeds, say, 3 or 4). It seems reasonable to expect, therefore, that phase correlation can provide a useful estimate of $\mathbf{s}_{m+1,m}$ $(= \mathbf{s}_{m+1} - \mathbf{s}_m)$ for some arbitrarily translating and slowly rotating objects. Fig. 4.14 and Fig. 4.15 (see §4.4.2) are computer simulated examples illustrating the sensitivity of phase correlation to, respectively, multiple exposure and rotation. It has been my experience that the phase correlation technique can tolerate, respectively, a lag factor less than 0.6 and a rotation of less then 5°. If multiple exposure and rotation occur simultaneously, phase correlation can be expected to be more restricted in its usefulness.

Note that when the motion is purely translational (i.e. there is no rotation, implying that each $\mathbf{R}_m$ is the unit dyadic), and m is large enough for the final term on the right hand side of (4.15) to be negligible, (4.15) reduces to

$$\begin{aligned} F_{m+1}(\mathbf{u}) &= F_m(\mathbf{u}) \exp(i2\pi \mathbf{u} \cdot [\mathbf{s}_{m+1} - \mathbf{s}_m]) \\ &+ \sum_{n=1}^{m} k^n A(\mathbf{u}) [\exp(i2\pi \mathbf{u} \cdot \mathbf{s}_{m+1-n}) \\ &\quad - \exp(i2\pi \mathbf{u} \cdot [\mathbf{s}_{m-n} + \mathbf{s}_{m+1} - \mathbf{s}_m])] \end{aligned} \tag{4.16}$$

In the particular case of uniform translational motion, for which

$$\mathbf{s}_m = m\mathbf{s} \tag{4.17}$$

where s is a constant shift, the summation in (4.16) vanishes. It follows that the capability of phase correlation to estimate the shift between successive CA images of a uniformly translating object is unaffected by the images being multiple exposures.

4.3 Practical Implementation

There are two main kinds of imperfection which need to be considered in order to usefully implement SPPC for video image sequences, namely *multiple object motion* and the types of noise (see §1.2.1 and §6.2.4) which commonly contaminate images.

The multiple object motion described in §1.2.1 is found to affect the implementation of PC more significantly than does the noise existing in X-ray coronary angiograms. The reason is that, due to the complicated motion of the heart, the coronary arteries appear deformable and overlapping. Heart muscle as well as all other parts of the body surrounding the heart form a non-stationary and non-uniform background. That is, the background contains significant intensity variations and often moves differently from any FOI (of interest to physicians). However, within a limited AOI, the FOI is often seen as the most distinct feature, because the angiographic system has been positioned appropriately for this purpose. Within an appropriately chosen AOI (typically 64 × 64 pixels), it can be assumed that the FOI moves as a rigid object on an featureless but non-uniform background. The presence of the background can reduce the effectiveness of PC. Therefore, steps are taken to reduce the background before invoking PC.

SEE ERRATA

For practical computational reasons the AOI is chosen to be rectangular, with its sides parallel to those of the camera target screen. It is therefore clear that any AOI is a truncated image. The effect of truncation (discussed in §2.2.2) must be reduced before applying PC. Image noise (see §1.2.1 and §6.2.4) is, of course, unavoidable in practice. Phase correlation is adversely affected by noise, so noise filtering is necessary for the success of the entire processing.

Three modifications to the idealised PC (as described in §2.3.4) are described in the following three sub-sections. §4.3.1 presents an algorithm compensating for the non-uniform background. §4.3.2 discusses compensating for truncation by invoking a Hanning window, which is useful for smoothing discontinuities at boundaries of an image. §4.3.3 demonstrates that applying a Blackman window in the frequency domain is very effective as a low-pass filter to reduce the influence of noise on phase correlation.

4.3.1 Compensating for Background

A general algorithm for removing backgrounds cannot be devised since the characteristics of the background are likely to be different for images obtained in different practical situations. The following discussion therefore mainly concerns coronary angiograms.

The background of a CA image tends to manifest itself as a relatively smooth featureless variation across an AOI, approximating a linear gradient in both the x and the y directions. An example is shown in Fig. 4.4. The background within the AOI shown in Fig. 4.4(a) is largely the light to dark transition from left to right. A section of the AOI in Fig. 4.4(a), as plotted in Fig. 4.4(b), which is representative of all horizontal sections through the AOI, is dominated by an approximately linearly varying background.

Phase correlation is adversely affected by the presence of such a linearly varying background. An algorithm is therefore invoked to subtract from each image $a_m(\mathbf{x})$ (which is formed by lag correction of the CA image $f_m(\mathbf{x})$, as described in §4.1.2) a linear function $\lambda(\mathbf{x})$ existing only within the AOI, i.e.

$$\lambda(\mathbf{x}) = \gamma_x x + \gamma_y y + \kappa \tag{4.18}$$

where γ_x, γ_y and κ are estimated by linear regression (least square fitting) of the function $a_m(\mathbf{x})$ within its AOI. Although this algorithm is simple, it has been found to be very effective for reducing the background. In addition to improving the performance of PC this linear background subtraction also leads to a significant improvement in the visibility of the FOI in the AOI. For example, Fig. 4.4(c) shows the image of the AOI (the boundary of which is highlighted in Fig. 4.4(a)) after subtraction of the linearly varying background. Before its removal, the background overshadows the FOI because the FOI occupies only a small portion of the dynamic range of the display device. Note further that this particular subtraction process does not add to the noise level of the image. Fig. 4.4(c) illustrates that the signal-to-noise ratio (SNR, which is defined in Chapter 6 for X-ray images) is unchanged after removing the linearly varying background.

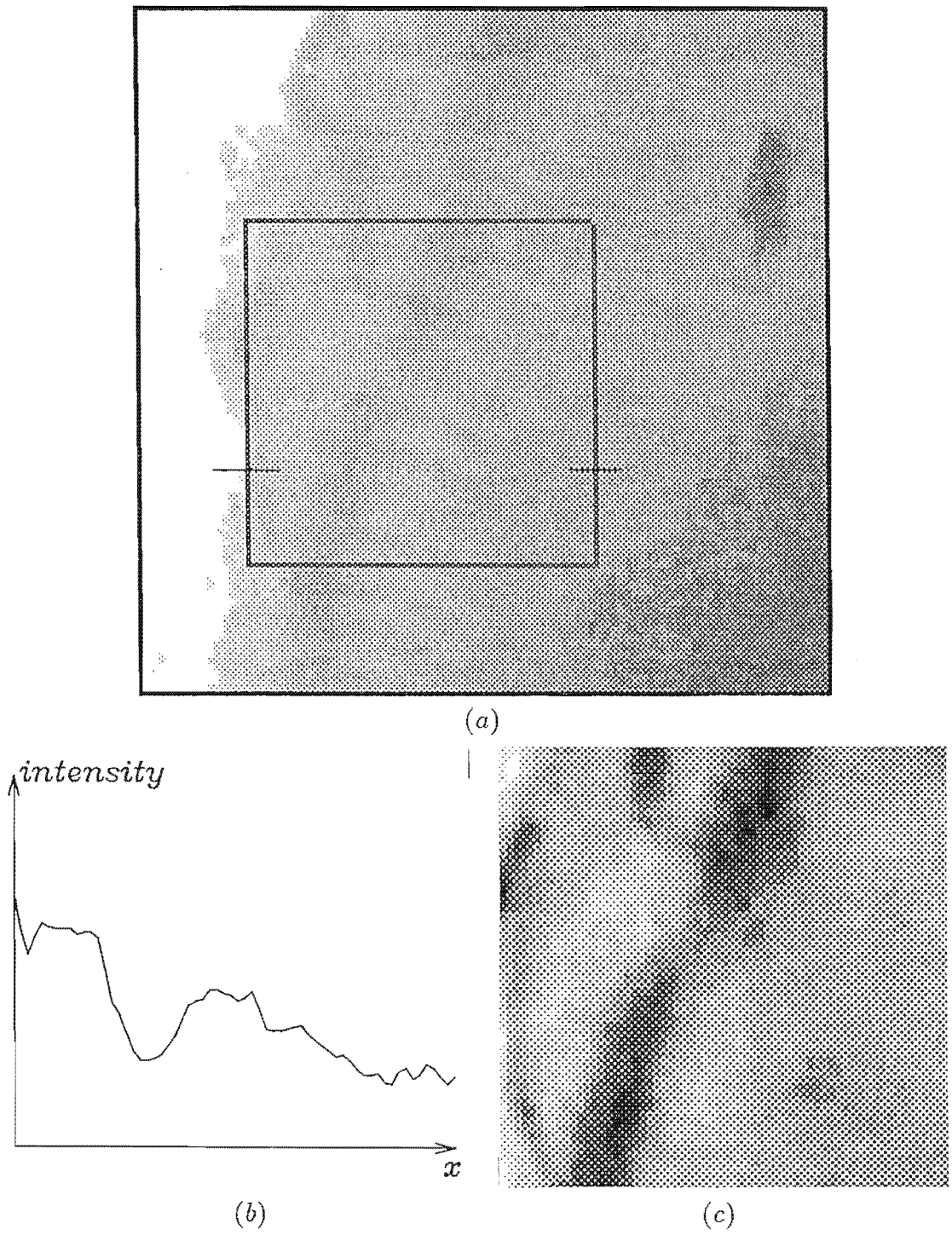


Figure 4.4. (a) typical CA image showing the right coronary artery with a significant narrowing within the AOI (the boundary of which is highlighted); (b) a horizontal section across the AOI, as indicated in (a); (c) image within the AOI after subtraction of linearly varying background. Note that in (c) the lag is not corrected before the subtraction of the background so that the blur (in the vertical direction) can still be noticed even though the FOI is enhanced by the background subtraction.

4.3.2 Ameliorating Effects of Image Truncation

The truncation effect and two general approaches for reducing it have been briefly discussed in §2.2.1. Some aspects of the truncation effect and how to remedy it in the context of the PC technique are disscussed in this sub-section.

In each image of an ideal sequence, the AOI contains the whole of the FOI and no part of any other cardiac feature (which can in general be expected to be executing a different motion from the FOI). Parts of other features cannot be prevented in practice from appearing within the AOI in some images, nor can all of the FOI always be present within the AOI. So, significant features, both of interest and otherwise, are in general truncated by the rectangular perimeter of the AOI.

Before discussing the amelioration of the truncation effects, it is helpful to discuss why the direct application of phase correlation to some truncated images is successful (see §2.3.4). It is mentioned in §2.3.4 that Alliney and Morandi [1986] find that applying phase correlation directly to 2-D truncated images produces results which are superior to those obtained from applying it directly to 1-D truncated images. Their argument (see §2.3.4) partially answers the question posed here. An alternative explanation is offered below, and is illustrated by applying PC to, respectively, a highly featured (or textured) image (Fig. 4.5) and a less textured image (Fig. 4.6).

Fig. 4.5(a) shows an image of an eagle cartoon and Fig. 4.5(b) depicts the shifted version of (a). Figs. 4.6(a) and 4.6(b) are two computer-generated images of a segment of a simulated artery, the second is shifted with respect to the first. The phase difference and phase correlation are shown, respectively, in Figs. 4.5(c) and 4.5(d) for the eagle, and in Figs. 4.6(c) and 4.6(d) for the simulated artery. It can be seen that the phase difference for the eagle has a distinct structure. The structure is the consequence of displaying the principal value of a 2-D phase function which is predominantly linear. For example, if a 2-D phase function is exactly linear as given by (4.8), any section of the phase function through the origin of Fourier space has the shape shown in Fig. 4.7, and such a section is here called a *sawtooth*. For convenience, the term "sawtooth phase" is henceforth invoked to describe the structure of a 2-D phase difference function which is linear. The similarity of the phase difference for the eagle to a sawtooth phase indicates that the phase difference function is effectively linear. On the other hand, the phase difference for the "artery" is seen to be a corrupted sawtooth phase, implying that the linearity of the function is significantly distorted. As a consequence of the difference between these two phase functions, the phase correlation of the eagle example has a single distinct peak at the correct position of the shift vector while that of the artery has many peaks. These two examples demonstrate that the result of directly applying phase correlation on images of the eagle is superior to that of directly applying it to the images of the "artery". It is known that, from the concept of phase dominance described in §2.3.1, it is the phase content which determines the features (edges, lines, etc. see §1.2.3) of an image. Because relatively fewer features are contained in the image of the "artery", the artifacts due to the truncation may be comparable to or even dominate the "artery" information in the phase. Therefore the phase difference so generated is very much influenced by the truncation effect. However, for the highly textured image

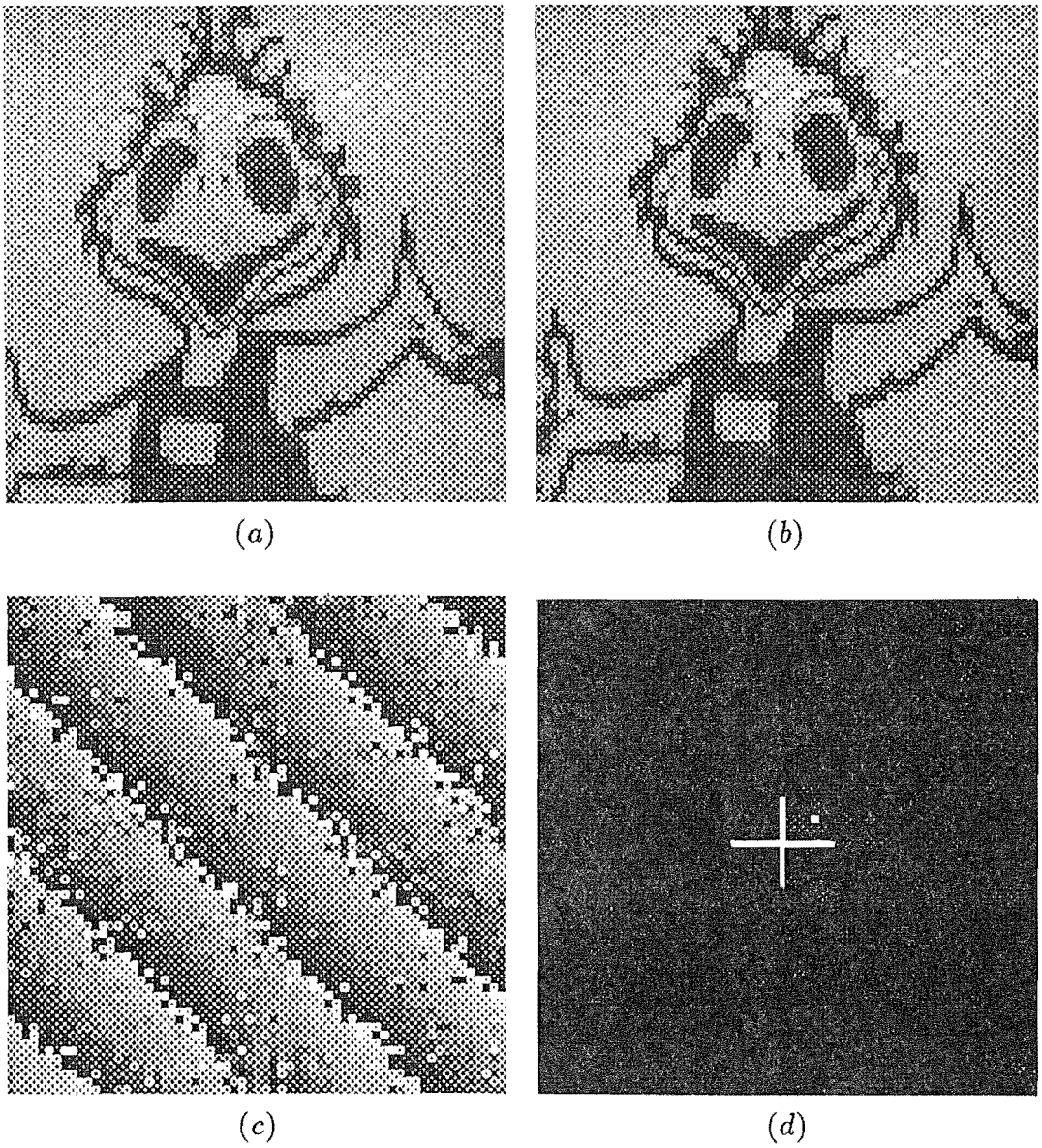


Figure 4.5. Phase correlation of two images of an eagle: (a) The original; (b) translated; (c) phase difference; (d) phase correlation. Note that the origin of phase correlation space in (d) is indicated with a white cross, and the position of the (unique) peak relative to the origin indicates correctly the shift of the eagle.

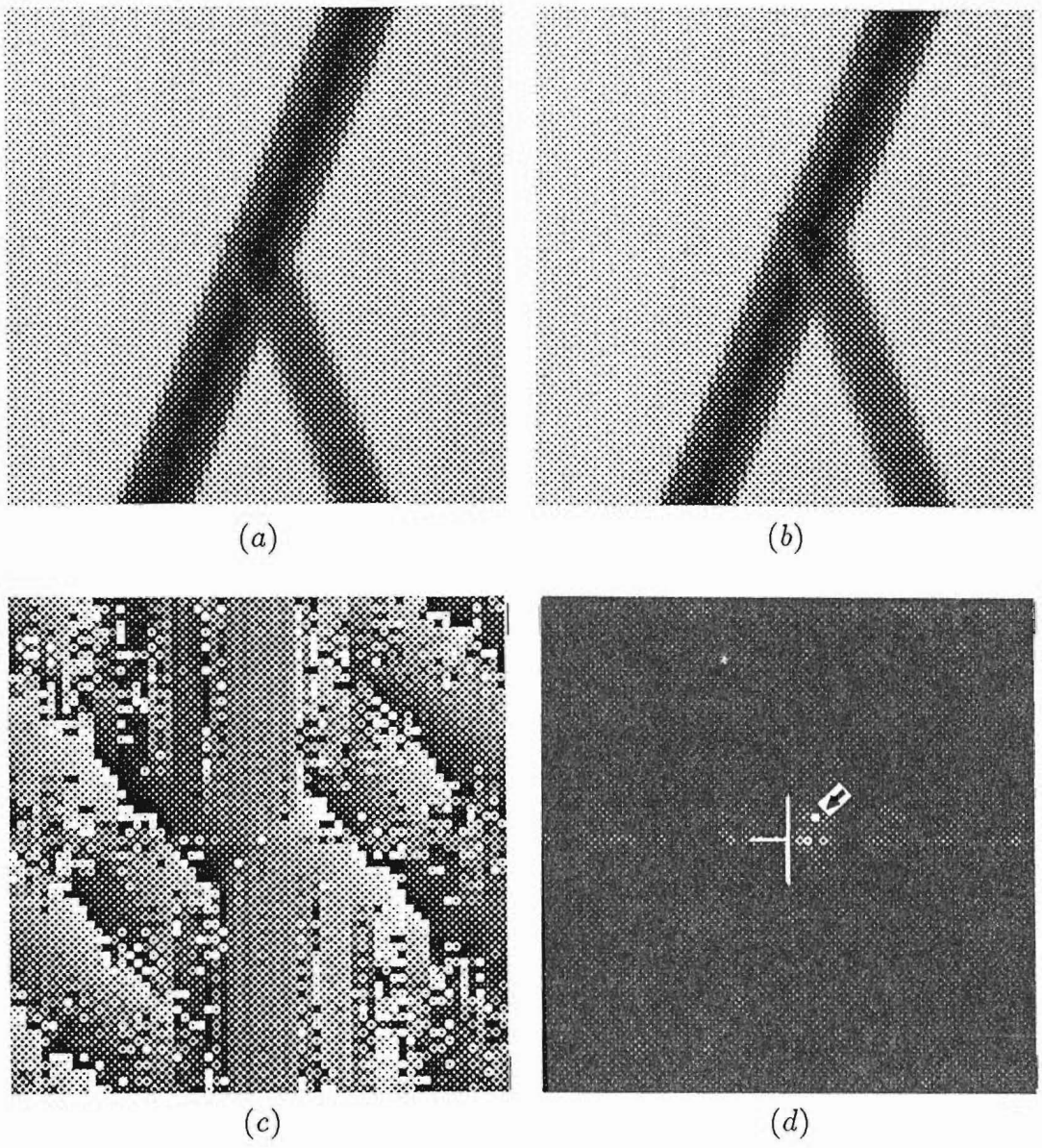


Figure 4.6. Phase correlation of two computer-generated images of a segment of a "coronary artery": (a) The original; (b) translated; (c) phase difference; (d) phase correlation. Note that the origin of phase correlation space in (d) is indicated with a white cross, and the position of the peak (indicated by an arrow) indicates the shift of the segment.

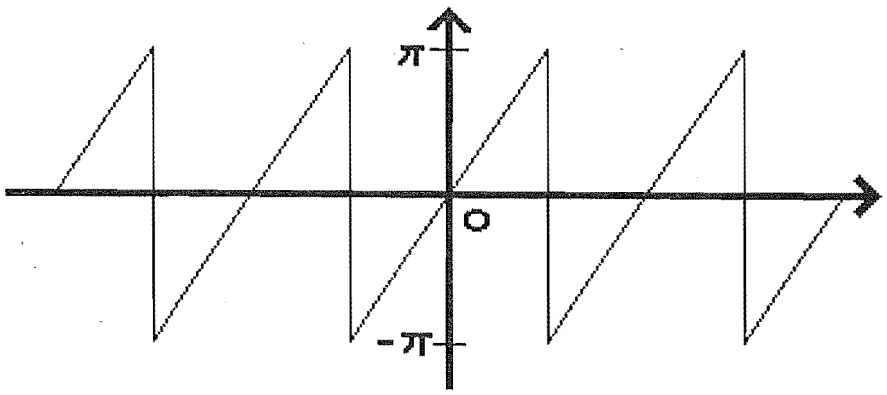


Figure 4.7. A typical profile through the principal value of a 2-D phase difference function which is exactly linear. Such a profile is here called a *sawtooth*, and the corresponding 2-D phase function is said to be a sawtooth phase.

of the eagle, although being truncated, the phase is still dominated by the image's feature information. This explains why the direct application of phase correlation gives different results for highly textured and less textured images.

As remarked in §2.2.2, "edge-extension" and "windowing" are two proven techniques for reducing the effects of truncation. While the former is often more efficacious, the latter tends to be more straightforward and is found adequate for angiographic processing because the effects of background and noise are usually at least as severe as those of truncation.

Harris [1978] comprehensively examines windows and their characteristics. More specifically, Alliney and Morandi [1986] consider, in the context of phase correlation, the use of symmetric windows which fall off smoothly to zero at the edges of image. Their justification for the use of symmetric windows is as follows. Given a pair of images of the same object, one being the shifted version of the other, and neither of them being truncated, the ratio of their respective Fourier transform magnitudes is unity, because of the shift invariant property of the Fourier transform magnitude. In the presence of the truncation effects, the ratio is generally not unity. Alliney and Morandi [1986] argue that a symmetric window is required to force the ratio to be as close to unity as possible.

It is emphasized that, while windowing ameliorates the consequences of truncation, it causes insignificant degradation to the phase information of the Fourier transform of the original images. This is again because the phase is related to features of the image rather than the absolute intensity distribution of the image. Windowing only puts a smooth weighting function on the overall intensity distribution, such that features (e.g. the edges of objects) remain in the image.

Before submitting CA images to phase correlation, I apply a Hanning window [Harris, 1978] to the contents of the AOI in each CA image from which the effect of camera lag has been removed in the manner indicated in §4.1.2. It preserves

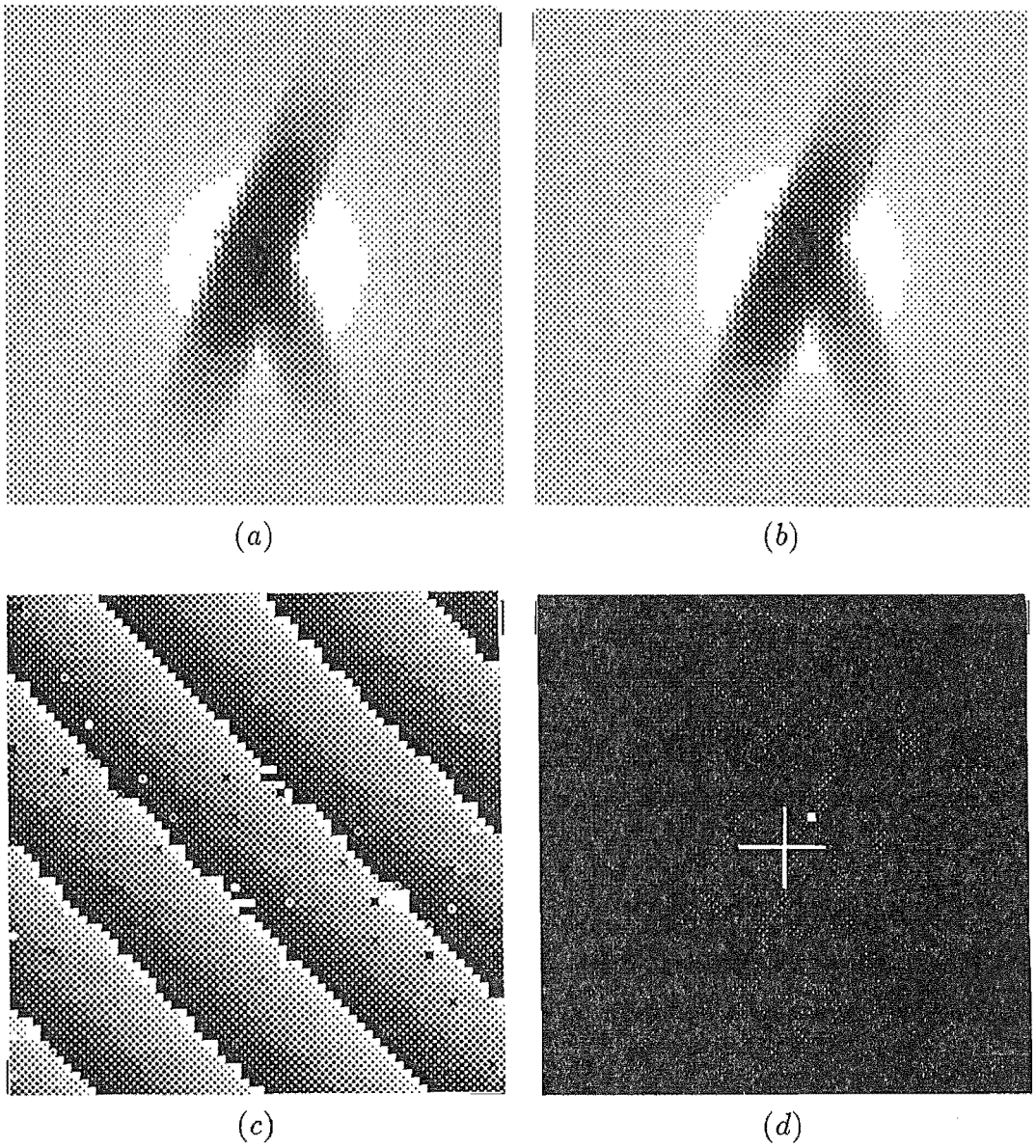


Figure 4.8. Effect of windowing on phase correlation: (a) and (b) Hanning windowed versions of Figs. 4.6(a) and 4.6(b); (c) phase difference between the images shown in (a) and (b); (d) phase correlation resulting from (c). Note that the origin of the phase correlation space in (d) is indicated with a white cross, and the position of the (unique) peak indicates correctly the shift of the vessel segment.

the phase information while effectively compensating for the truncation. Fig. 4.8 illustrates the usefulness of Hanning windowing on the computer-simulated "artery" shown in Fig. 4.6.

Figs. 4.8(a) and 4.8(b) show Hanning-windowed versions of Figs. 4.6(a) and 4.6(b) respectively. The phase difference shown in Fig. 4.8(c) clearly has the appearance

of a sawtooth phase, indicating that the linearity of the phase difference function is restored by the windowing (comparing with Fig. 4.6(c)). The phase correlation shown in Fig. 4.8(d) has a distinct peak at the position corresponding accurately to the relative shift between images shown in Figs 4.8(a) and (b) (compare with Fig. 4.5(d) and Fig. 4.6(d)).

4.3.3 Compensating for Noise

It is argued in §6.2.4 that the noise process in a digital angiographic system is Gaussian. Experience also shows that the noise has a frequency spectrum which is essentially "white", throughout the range of spatial frequencies characterising a coronary angiographic image. After the subtraction operation described in §4.1.2 has been invoked, the noise variance tends to increase as predicted by (4.4). In practice, the spatial frequency content of $A(\mathbf{u})$ (see §4.2) is limited, being effectively restricted to some finite part of Fourier space. The high frequency region is therefore likely to be dominated by the noise spectrum. So, a composite spectrum is formed by combining the phase difference with a suitable magnitude $W(\mathbf{u})$, rather than with unity magnitude as described in §4.2.1,

$$C_{m+1,m}(\mathbf{u}) = W(\mathbf{u})\exp(i\Phi_{m+1,m}(\mathbf{u})) \tag{4.19}$$

where $\Phi_{m+1,m}(\mathbf{u})$ is the phase difference between $A_{m+1}(\mathbf{u})$ and $A_m(\mathbf{u})$, and $W(\mathbf{u})$ is real and non-negative. Since the spectrum of a noise-free image of a typical FOI is concentrated at low frequencies (i.e. near the origin in Fourier space), $W(\mathbf{u})$ is chosen to be a symmetric "window" function having a value close to unity throughout most of the aforementioned part of Fourier space but falling smoothly to zero outside it. Defining $c_{m+1,m}(\mathbf{x}) \longleftrightarrow C_{m+1,m}(\mathbf{u})$, it is seen that $c_{m+1,m}(\mathbf{x})$ has the character of an impulse of finite width centered at $\mathbf{x} = \mathbf{s}_{m+1,m}$. Computational experience has led me to adopt the circularly symmetric Blackman window function [Harris, 1978] for $W(\mathbf{u})$.

It is emphasised that $c_{m+1,m}(\mathbf{x})$ is still more sharply peaked than the conventional cross-correlation, even though $W(\mathbf{u})$ is not everywhere unity. The point is that the effective width of $W(\mathbf{u})$ can be usefully chosen to be larger than that of the central lobe of $|A(\mathbf{u})|^2$. Because of the complementary natures of image space and Fourier space, illustrated by Fig. 4.9, the spatial extent of $c_{m+1,m}(\mathbf{x})$ is therefore smaller than that of the cross-correlation determined by the inversion of $|A(\mathbf{u})|^2$.

Fig. 4.10 illustrates the effect of Blackman filtering on processing noisy images by phase correlation. Figs. 4.10(a) and 4.10(b) are obtained by adding computer-simulated zero-mean Gaussian noise to Figs. 4.6(a) and 4.6(b) respectively. Fig. 4.10(c) shows the phase difference obtained by first (Hanning) windowing both Figs. 4.10(a) and 4.10(b) and then Fourier transforming them. Although very noisy, the central part of the phase difference still has the appearance of a sawtooth phase, indicating that the "signal" dominates the noise at low frequencies. The phase correlation obtained directly from Fig. 4.10(c) is shown in Fig. 4.10(d). The presence of the noise has produced a number of sharp peaks. The position of maximum intensity (indicated with a arrow) is no longer at the position indicating the relative shift as

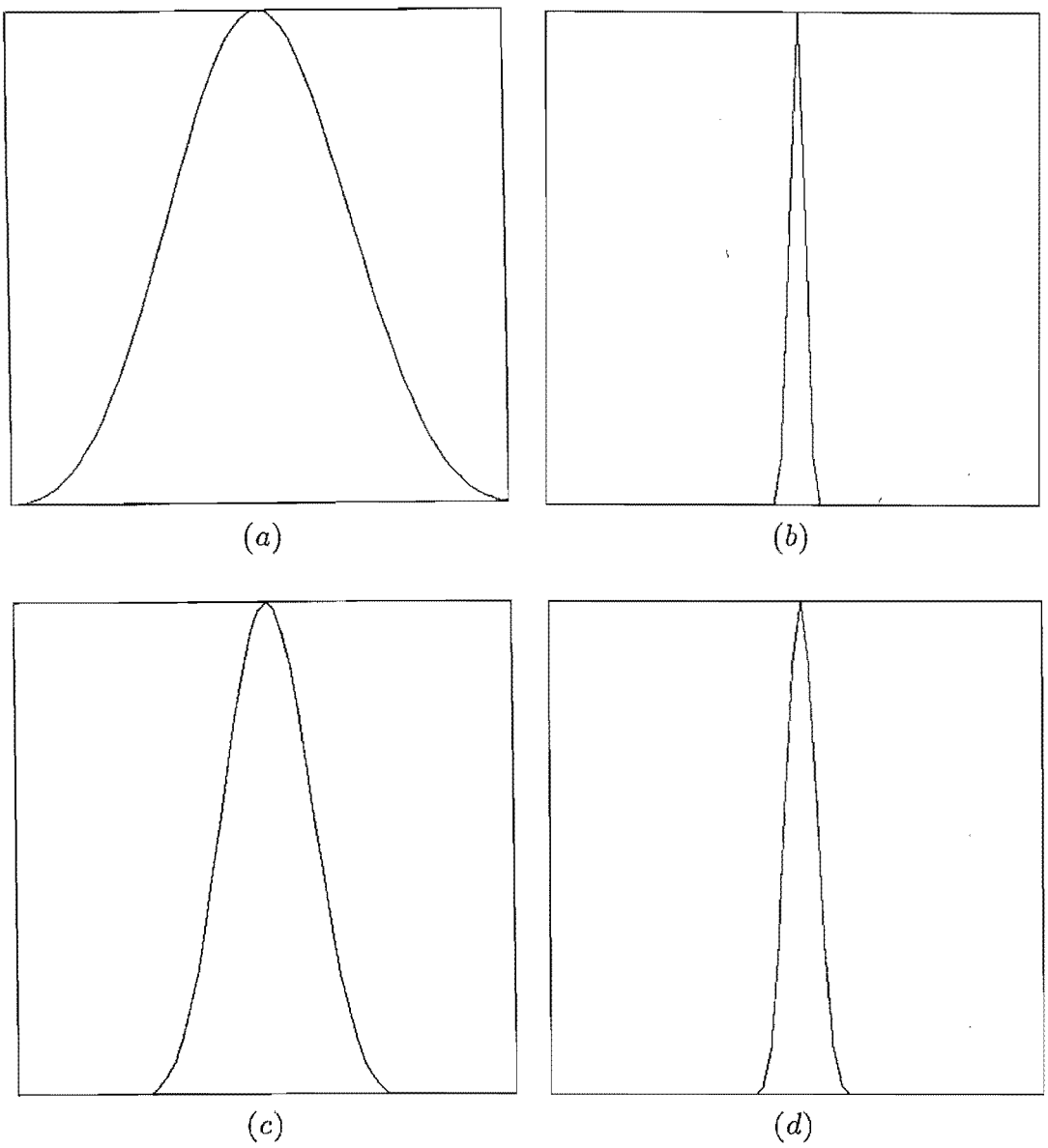
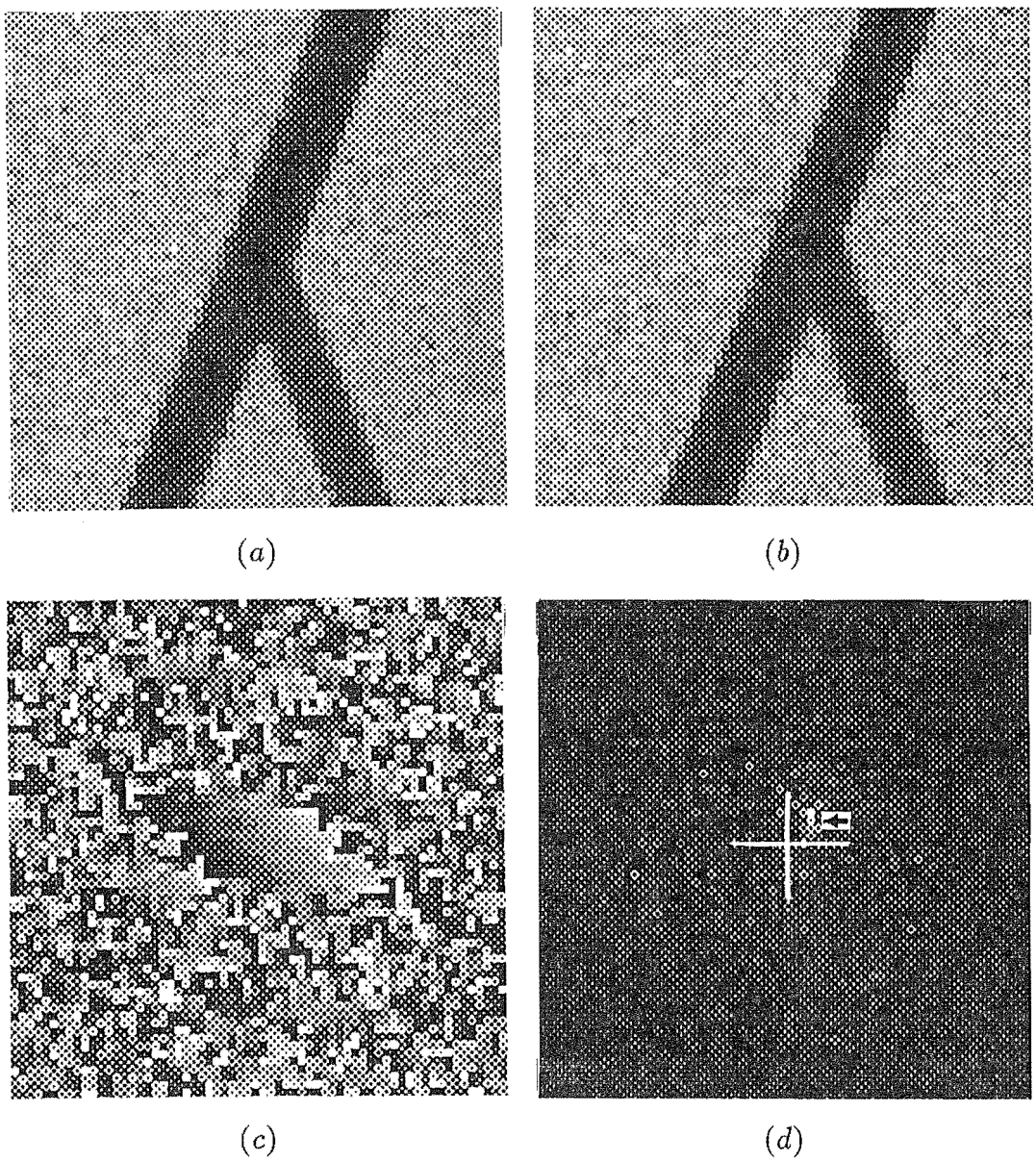


Figure 4.9. Illustration of the complementary natures of image space and Fourier space: (a) relatively broad one-dimensional Blackman window function; (b) Fourier transform magnitude of the window function shown in (a); (c) relatively narrow one-dimensional Blackman window; (d) Fourier transform magnitude of the window function shown in (c).



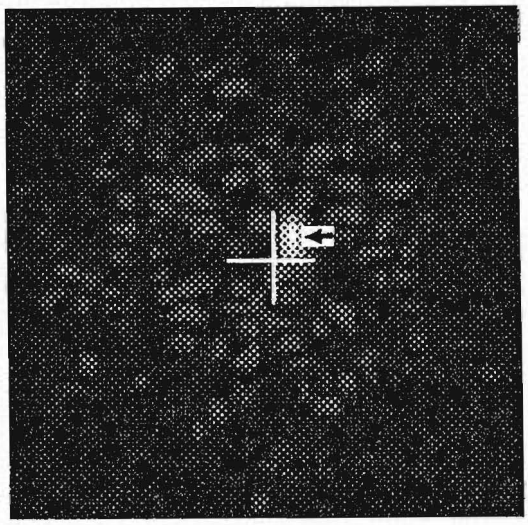
(a) (b)

(c) (d)

Figure continues on next page....

is confirmed by comparison with Fig. 4.8(d). When the phase difference is combined with the Blackman window (and is thereby "filtered"), however, the phase correlation shown in Fig. 4.10(e) results. The maximum peak is now correctly positioned although relatively broad in comparison with the peaks obtained without filtering.

It should be mentioned that other ways of ameliorating the effects of noise are possible. For instance, one could filter the images before processing them by phase correlation. In particular, the median filter [Justuson, 1981], which is suitable for reducing both salt-and-pepper and burst noise, can be expected to improve images obtained from the subtraction operation described in §4.1.2. Fig. 4.11 illustrates the



(e)

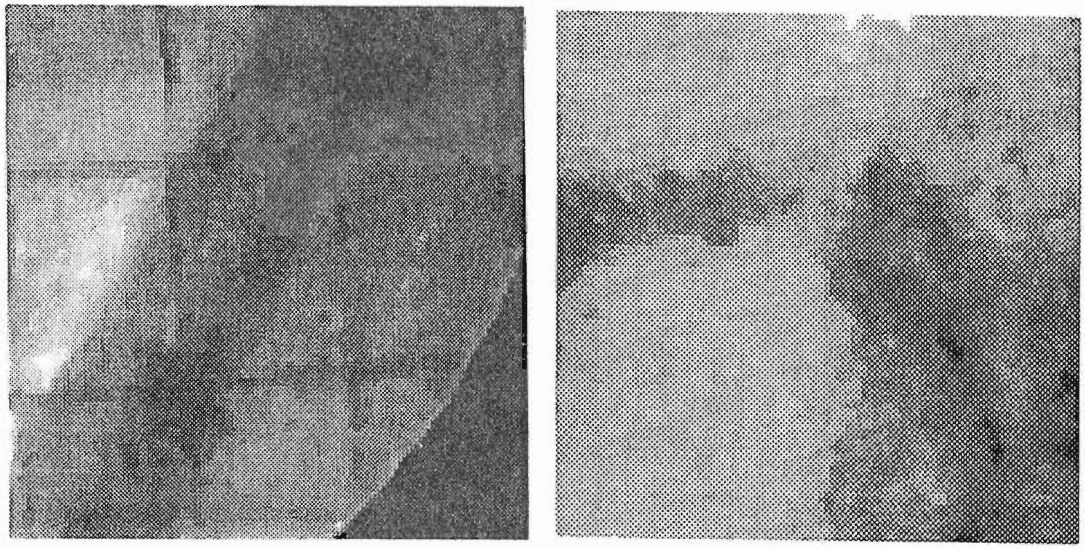
Figure 4.10. Effect of Blackman filtering on phase correlation of noisy images of "artery": (a) noisy version of the "artery" image shown in Fig. 4.6(a); (b) noisy version of the shifted "artery" image shown in Fig. 4.6(b); (c) phase difference between the images shown in (a) and (b); (d) phase correlation of the images shown in (a) and (b) without filtering; (e) phase correlation with Blackman filtering. Note that, in both (d) and (e), a white cross represents the origin, and a arrow points to a position which indicates the correct shift of the "artery" segment.

results of median filtering the lag corrected images shown in Figs. 4.2(d) and 4.3(d). Since the emphasis here is on the PC technique per se, such additional filtering is not discussed further.

Once the object motion is correctly estimated for an image sequence, averaging of all or several images after motion compensation can produce an enhanced image of the object. This is an important step in practice for many digital angiographic studies (see Chapter 6). Averaging of N images can result in a $\sqrt{N}$ improvement in SNR [Huang and Hsu, 1981]. Remembering also that the lag correction tends to amplify the noise inherent in digital angiographic images (see §4.1.2), the averaging process enables the SNR lost during lag correction to be recovered.

4.4 Algorithms and Illustrations

It is here appropriate to summarise what has already been demonstrated in this chapter. It is shown in §4.1 that weighted subtraction generates usefully faithful but noisy versions of the image of the FOI. Based on the phase correlation technique described in §2.3.4, §4.2 introduces a new, although still idealised technique, the SPPC, which is appropriate for processing image sequences. Three modifications to phase correlation, taking account of image imperfections encountered in practice,



(a) (b)

Figure 4.11. Illustration of median filtering: (a) median filtered version of Fig. 4.2(d); (b) median filtered version of Fig. 4.3(d).

are introduced in §4.3. These enable SPPC to process efficaciously image sequences having real world imperfections. Fig. 4.12 illustrates the three modifications. Note that in Fig. 4.12, $a_m(\mathbf{x})$ and $a_{m+1}(\mathbf{x})$ represent any two consecutive images in a sequence. The RS of these two images, $\mathbf{s}_{m+1,m}$, is estimated by the modified phase correlation technique.

The actual algorithms invoked for processing image sequences are described and illustrated in this section. The algorithms estimate the motion parameters (e.g. the shifts) of an FOI throughout an image sequence and obtain an enhanced image of the FOI by averaging after shift compensation.

Note that §4.4.1 below presents an algorithm which performs all the processing summarized above. Furthermore, three variants of the algorithm are introduced for completeness. Illustrations of the use of these algorithms on computer-generated images are presented in §4.4.2. Results of applying these algorithms to actual CA image sequences are reported in Chapter 7.

4.4.1 A Practical Image Sequence Processing Algorithm with Variants

The given sequence of images (containing some number, here denoted by M, of individual images) is processed by the following algorithm. The result of each processing step in the algorithm is indicated by the item to the right of the symbol "⇒". In addition, a block diagram of the algorithm is shown in Fig. 4.13.

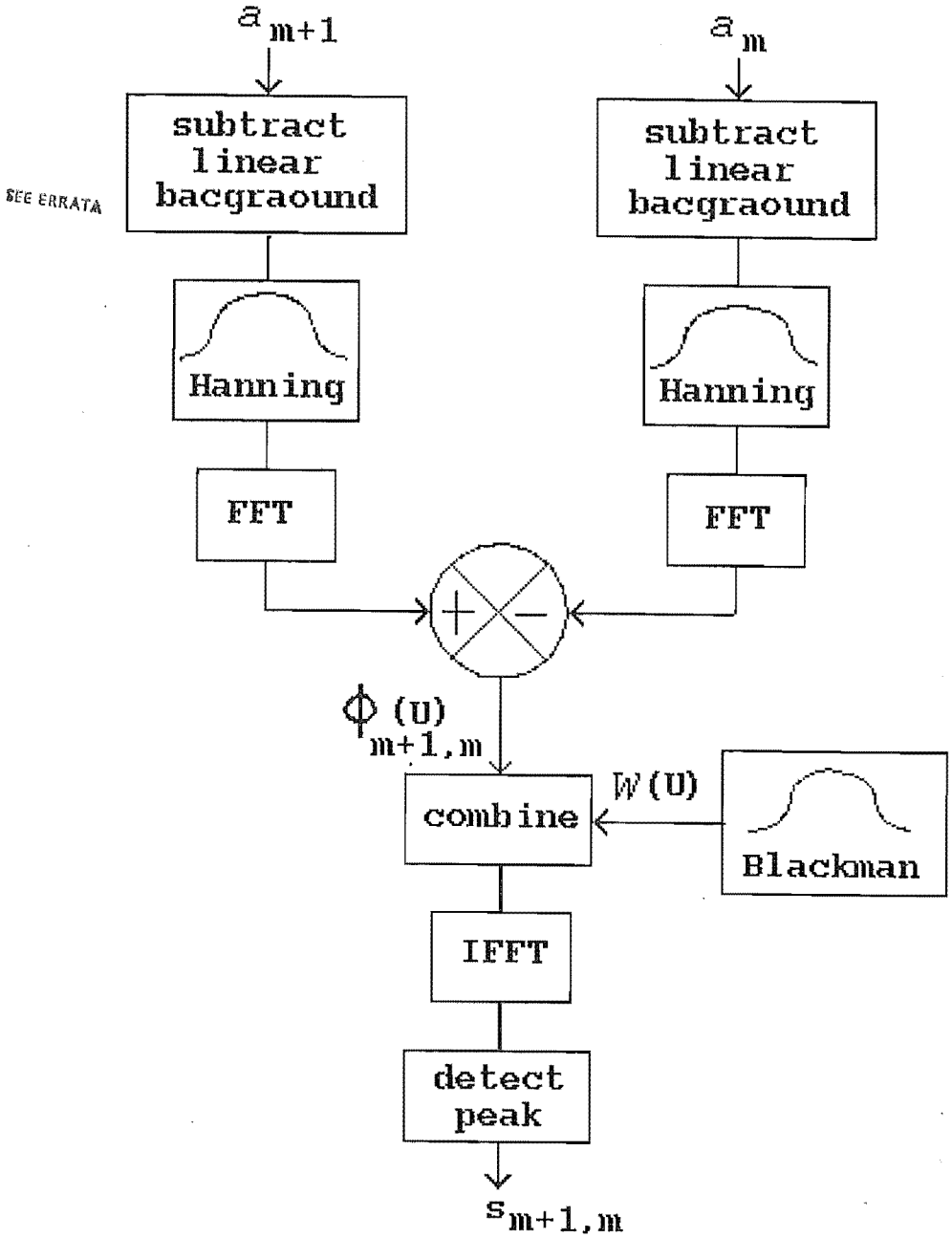


Figure 4.12. Block diagram illustrating the three modifications to the phase correlation technique, i.e. compensating for background, truncation, and noise, as discussed in §4.3.

Algorithm

1. Perform the subtraction operation on the image sequence $f_m, m = 1 \cdots M$, as prescribed by (4.3), compensating for multiple exposures $\Rightarrow a_m, m = 1 \cdots M-1$.

2. Initialize (process the first frame) by:

   Begin

   Choose the AOI to encompass the FOI from the first frame a_1 at a position $\mathbf{P}_1$, thus forming a truncated image also denoted, for convenience, by a_1.

   Remove the linear background from a_1 (refer to (4.18)) $\Rightarrow a_1^{(1)}$.

   Apply a Hanning window to $a_1^{(1)} \Rightarrow a_1^{(2)}$.

   Let $\mathbf{s}_{1,0} = \mathbf{s}_1 = (0,0)$, (refer to (4.11)).

   End;

3. Process the remaining frames in a loop:

   For $m = 2$ To $M - 1$ Do

   Begin

   Let $\mathbf{P}_m = \mathbf{P}_{m-1} + \mathbf{s}_{m-1,m-2}$;

   Extract the AOI from a_m, also denote it by a_m;

   Subtract the linear background from a_m (refer to (4.18)) $\Rightarrow a_m^{(1)}$;

   Window $a_m^{(1)}$ with Hanning function $\Rightarrow a_m^{(2)}$;

   Perform phase correlation for images $a_m^{(2)}$ and $a_{m-1}^{(2)}$ to obtain the shift vector $\Rightarrow \hat{\mathbf{s}}_{m,m-1}$;

   Calculate the actual relative shift $\Rightarrow \mathbf{s}_{m,m-1} = \hat{\mathbf{s}}_{m,m-1} + (\mathbf{P}_m - \mathbf{P}_{m-1})$;

   End;

4. Reposition the AOI in $a_m, m = 2 \cdots M - 1$, by compensating for each relative shift $\mathbf{s}_{m+1}, m = 2 \cdots M - 1$. (refer to the third row in Fig. 4.13).

5. Superimpose the content of the new AOI from $a_m, m = 1 \cdots (M-1)$ to generate a single image $\tilde{a}$, the enhanced AOI. (refer to the fourth row in Fig. 4.13).

Variants of this algorithm can be obtained by modifying it in at least three aspects:

1. Estimate AS rather than RS if the deformation of the FOI (see §4.2.1) can be ignored throughout the sequence (with the advantage of avoiding the cumulative error of calculating an AS from all previous RS's);

2. Apply phase correlation to original video images rather than to lag-compensated images when k is small, as justified in §4.2.2;

3. Apply median filtering after the lag removal step in the algorithm (it is found to be a useful step for processing images with a high noise level).

Chapter 7 presents and discusses results of applying algorithms presented in this sub-section to clinical CA image sequences.

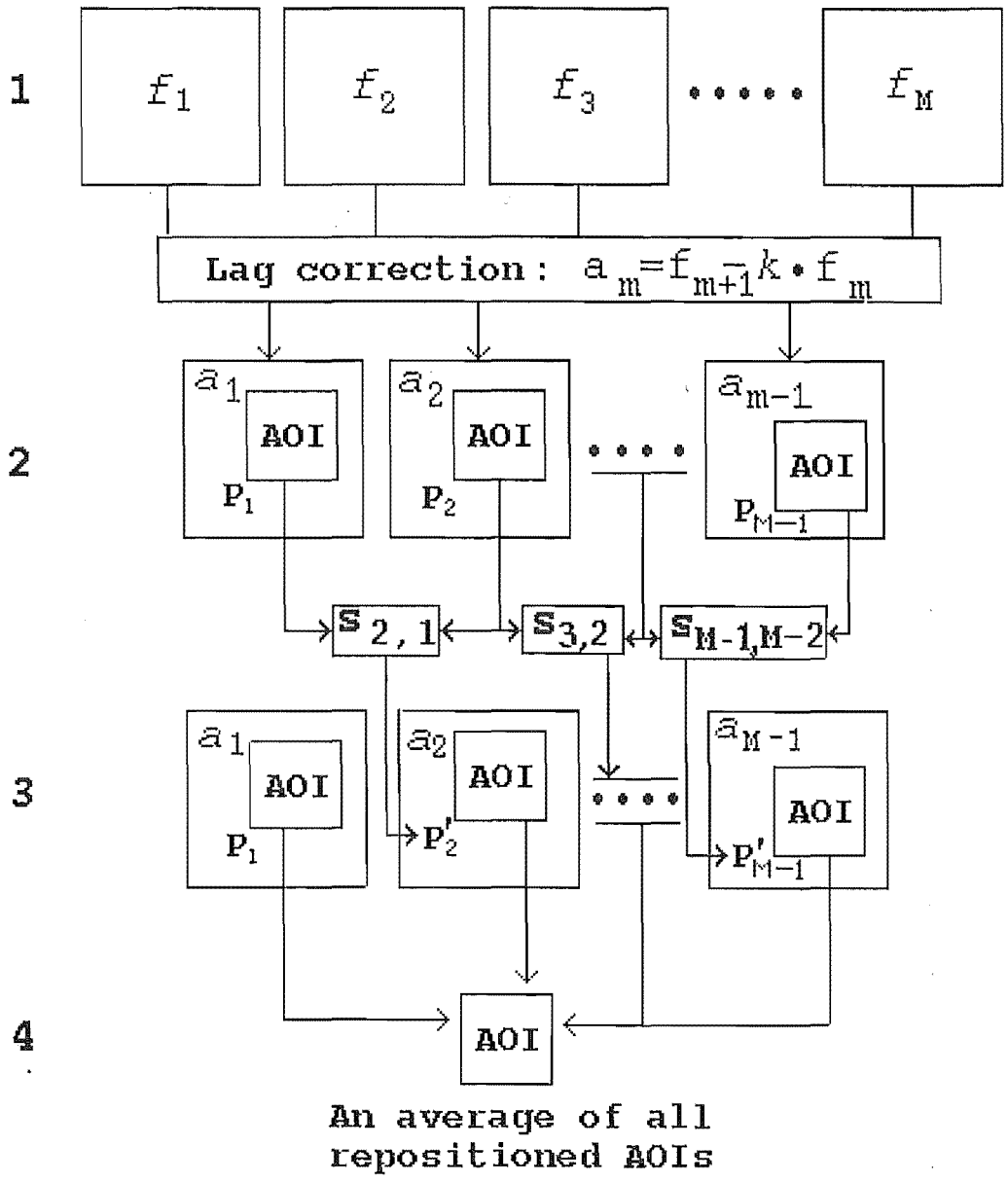


Figure 4.13. Block diagram of a practical image processing algorithm for estimating the translational motion of a FOI by SPPC, and superimposing images for enhancement of the FOI. The original CA image sequence, $f_m, m = 1 \cdots M$, is "displayed" in the first row. The lag correction produces $M-1$ images, $a_m, m = 1 \cdots M-1$, "shown" in the second row. The SPPC is then performed to estimate relative shifts in the image sequence (Note that the initial AOI is first positioned inside a_1 and a_2 SEE ERRATA for estimating RS $s_{2,1}$ which in turn is used to position the AIO in a_3). After all $(M-2)$ relative shifts $s_{m+1,m}, m = 1 \cdots M-2$ are estimated, they are used to reposition AOIs in $a_m, m = 2 \cdots M-1$, as "displayed" in the third row. Finally, the content of the new AOI in all a_m are superimposed for obtaining a averaged image of the AOI as "shown" in the fourth row.

4.4.2 Illustrative Examples

I have simulated the X-ray shadow image of a coronary artery (or, more specifically, the contrast medium within the lumen of the artery) by calculating the projection of two intersecting homogeneous circular cylinders according to the exponential law of attenuation (see §6.2). The cylinders are taken to be embedded in a homogeneous region of lower attenuation.

I have noticed that the phase differences can be more informative than the phase correlations themselves. Often, by examining the phase difference, the usefulness of phase correlation can be predicted to some degree. Therefore, illustrations given in previous and current sections mostly include phase differences.

Figs. 4.5, 4.6, 4.8 and 4.10, described in previous sections of this chapter, are results of computer simulations which illustrate the effectiveness of the practical implementation of the phase correlation technique as presented in §4.3. In what follows, two additional examples of computer-simulation are given in Fig. 4.14 and Fig. 4.15, both of which are referred to in §4.2.2, illustrating the sensitivities of PC to multiple exposure and rotation respectively.

In Fig. 4.14(a) and (b), the f_m and f_{m+1} (cf. §4.2.2), are obtained in the following three steps: firstly, 11 images, here denoted as $a_j, j = 1 \cdots 11$, are generated by translating the "coronary artery" shown in Fig. 4.6(a) sinusoidally in the horizontal direction, with maximum shift being 10 pixels from its original position; secondly, computer-simulated zero-mean Gaussian noise is added to each a_j to produce a_j^N, where the SNR is of similar order of magnitude to that for actual CA images (see Fig. 4.2(a) or Fig. 4.3(a)) (note that the noise added to each of the eleven images is statistically independent of the others); thirdly, f_m and f_{m+1} are produced by, respectively,

$$f_m = f_{10} = k^{10}a_1^N + k^9 a_2^N + \cdots + a_{10}^N \tag{4.20}$$

and

$$f_{m+1} = f_{11} = k^{10}a_2^N + k^9 a_3^N + \cdots + a_{11}^N \tag{4.21}$$

where the persistence factor $k = 0.55$ (so the influence of previous frames from a_{m-11}^N can be neglected).

Fig. 4.14(c) is the phase difference between the Fourier transforms of Hanning windowed versions of f_m and f_{m+1}. It still has the appearance of a sawtooth phase in its central part. Fig. 4.14(d) is the phase correlation obtained by Blackman filtering the phase difference spectrum (see §4.3.3), followed by performing the inverse Fourier transformation. The peak (indicated by an arrow) in Fig. 4.14(d) is correctly positioned, representing the shift between a_{10}^N and a_{11}^N, i.e. $a(\mathbf{x} + \mathbf{s}_{11})$ and $a(\mathbf{x} + \mathbf{s}_{10})$. The result indicates that the presence of significant video camera lag does not prevent an accurate detection of shift (as argued in §4.2.2) even in the presence of significant random noise.

In order to study the sensitivity of phase correlation to rotation, a sequence of 8 computer-generated images of an artery segment which rotates gradually is studied. The first of the sequence is chosen as the reference and the rest have the same shift (3 pixels upward and 3 pixels rightward) with respect to the reference, but rotated

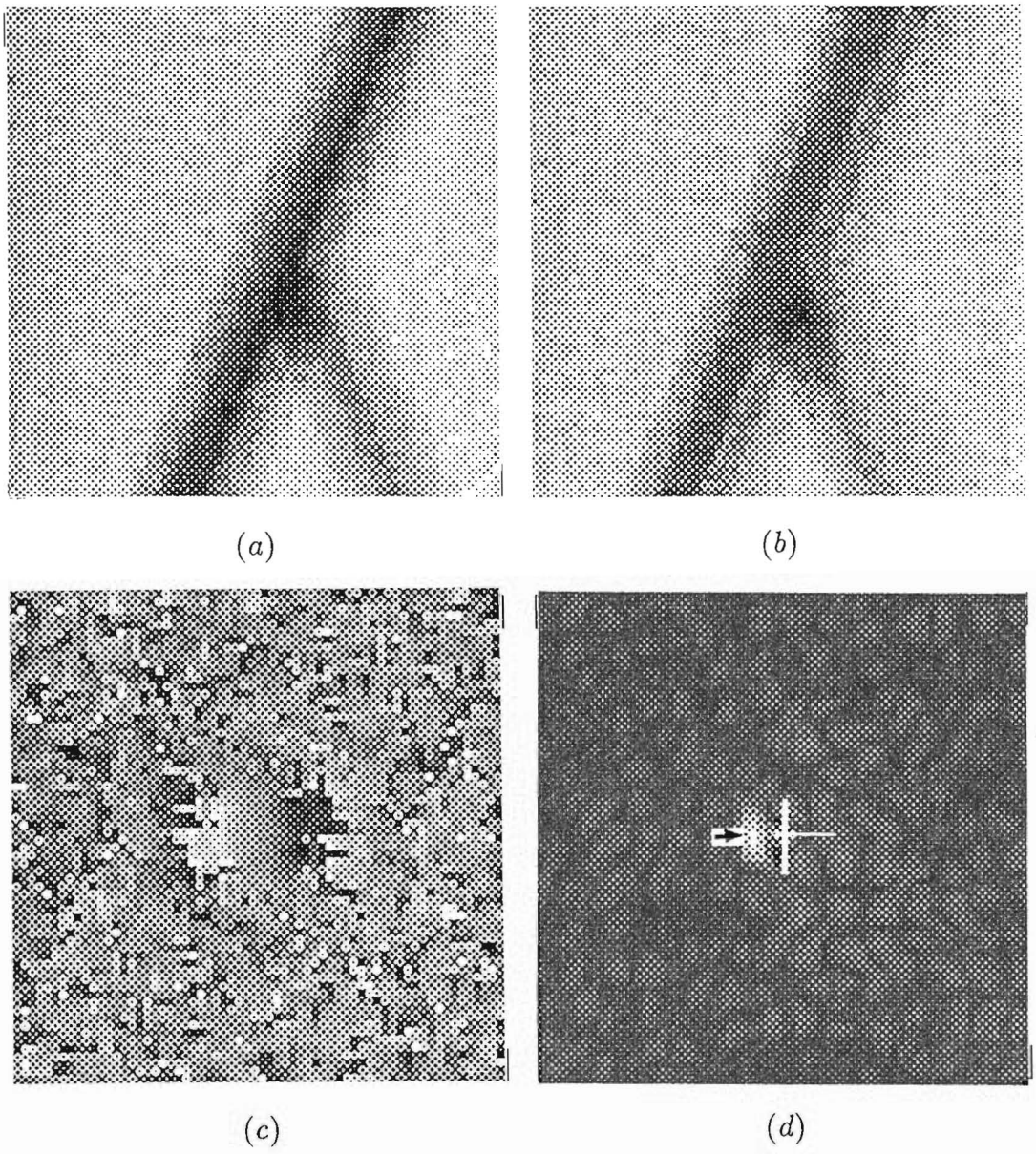
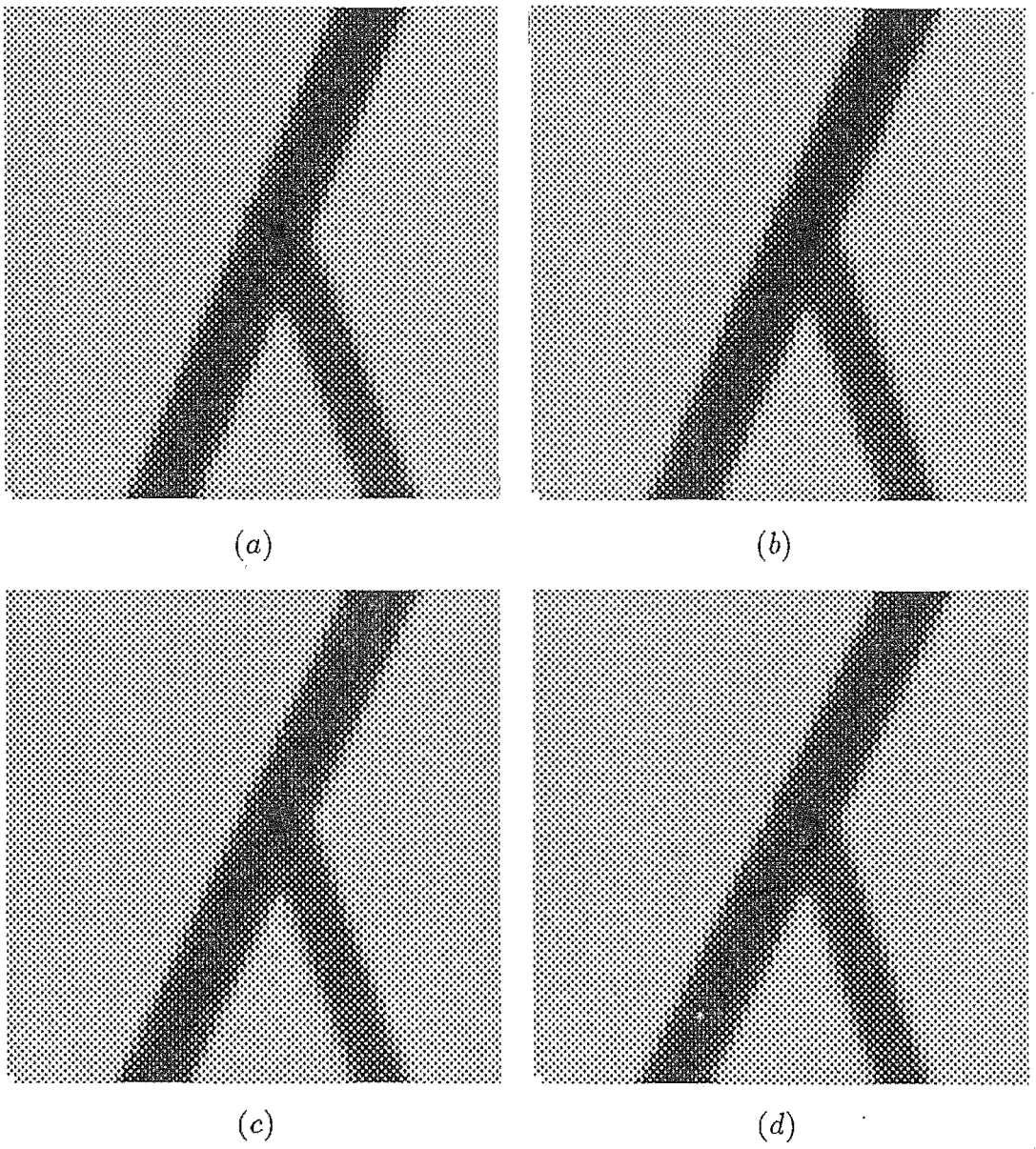


Figure 4.14. Phase correlation of images being noisy and multiple exposures: (a) and (b) are consecutive images multiply exposed according to, respectively Equation 4.20 and 4.21 ($k = 0.55$), with non-uniform horizontal translation simulated (the x component of s_{m-n} varied sinusoidally); (c) phase difference between the Fourier transforms of (a) and (b); (d) phase correlation, with the peak (indicated by an arrow) representing the correct RS between a_{11}^N and a_{10}^N.

$1°, 2°, 3°, ...8°$ and so on. Part of the results (sufficient to illustrate the effect) are shown in Fig. 4.15. The original reference image is shown in Fig. 4.6(a). Fig. 4.15(a), (b), (c) and (d) are the translated and rotated versions (where the rotated angles are, respectively, $\theta = 3°$ for (a), $\theta = 4°$ for (b), $\theta = 5°$ for (c) and $\theta = 6°$ for (d)) of the reference image. Phase correlation performed for Fig. 4.6(a) and each of the images displayed in Fig. 4.15(a), (b), (c) and (d) results in, correspondingly, phase differences shown in, respectively, Fig. 4.15(e), (f), (g) and (h). Gradual stretching and distortion of these sawtooth phases can be seen from Fig. 4.15(e), (f), (g) and (h) as the rotation increases from 3° through 6°. After Blackman filtering followed by the inverse Fourier transformation of the phase difference spectra corresponding to (e), (f), (g) and (h), the correct translations are detected for cases (a) and (b) in Fig. 4.15, but not for (c) and (d), i.e. up to 4° rotation. The estimated shifts for (c) and (d) are both in error of 1 pixel horizontally.

The topics presented in this chapter are summarized in Chapter 8. The significance of these topics is also assessed in that chapter.



(a) (b)

(c) (d)

Figure continues on next page....

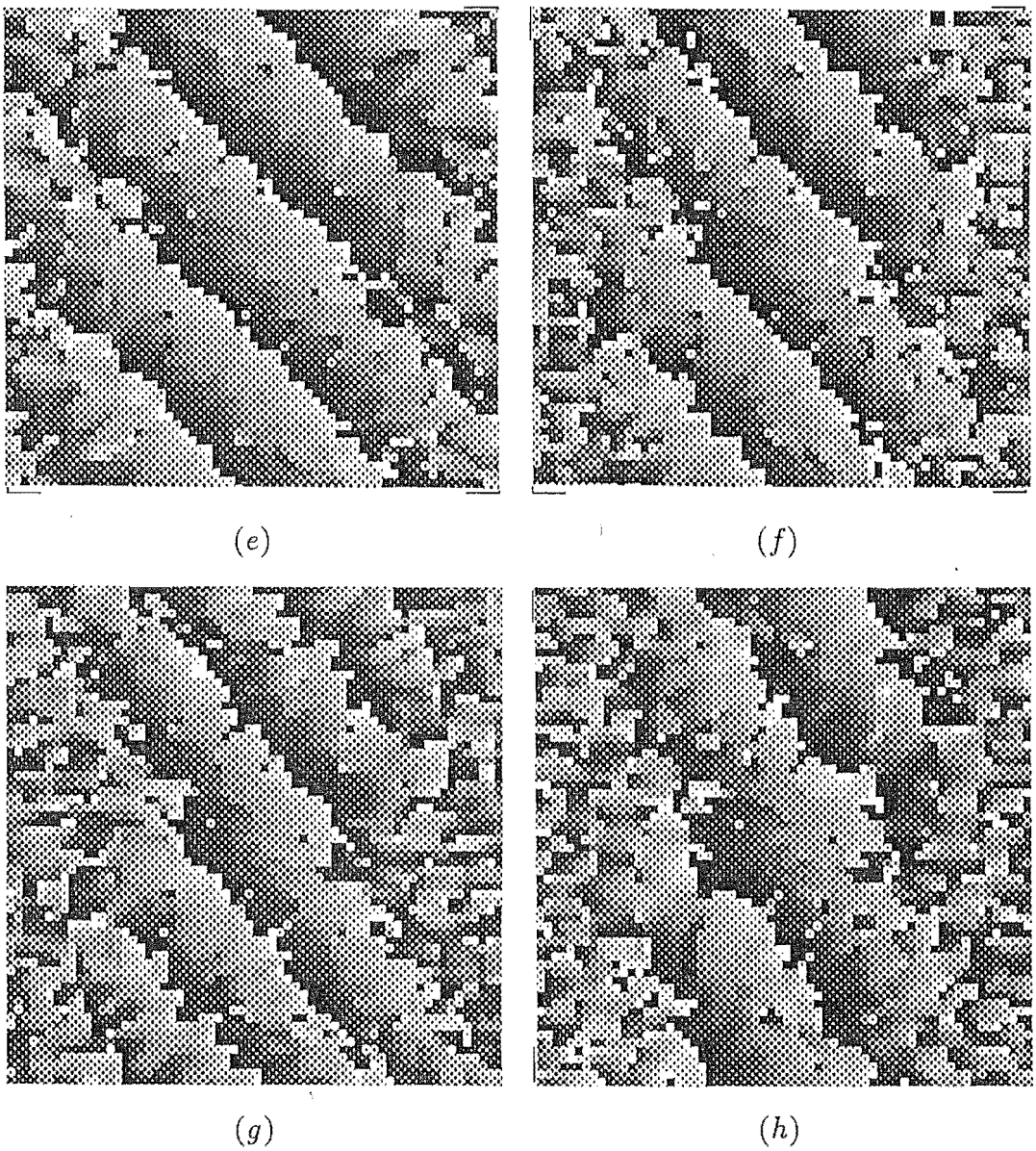


Figure 4.15. The effect of rotation on phase correlation of translation detection. Noise-free computer-generated images with identical translation but the following increasing rotations with respect to Fig. 4.6(a) are shown: (a) 3°; (b) 4°; (c) 5°; (d) 6°. Phase differences computed for Fig. 4.6(a) and each of (a), (b), (c) and (d) are shown correspondingly in (e),(f),(g),(h). In this example inverse Fourier transformation resulted in correct translation being detected in (a) and (b) only, i.e. with up to 4° rotation.

Chapter 5

Implementation of Angular Correlations for Rotation Estimation

The spectral and bispectral (angular) correlations are established respectively, in §2.3.6 and §3.3 for rotation estimation. In this chapter, aspects of implementing these two correlation functions are discussed. Like spectral angular correlation, bispectral angular correlation can also be invoked as part of the general 2-D motion estimation method discussed in §2.3.6, i.e. first estimating rotation with the aid of shift-invariant transforms, then, after compensating for rotation, estimating the translation.

The aim in this chapter is to compare the basic capabilities of the two correlation techniques. So as to avoid the possibility of confusing the outcome of the comparison with extraneous considerations, only binary images are considered. These images are enclosed in frames which include the complete image of whatever object is of interest (i.e. there is no truncation of the object of interest). Pseudo-random noise is added to frames (i.e. the images are contaminated).

The images invoked in the examples presented in this chapter are introduced in §5.1. Rotation estimation by ordinary spectral and bispectral angular correlations is discussed in §5.2. Next, in §5.3, the levels of performance to be expected of the two angular correlation techniques are illustrated and are compared with the aid of computer generated examples. Finally, in §5.4, the sensitivities of angular correlation functions and the computational efficiencies of the algorithms presented in this chapter are discussed.

5.1 Illustrative Images

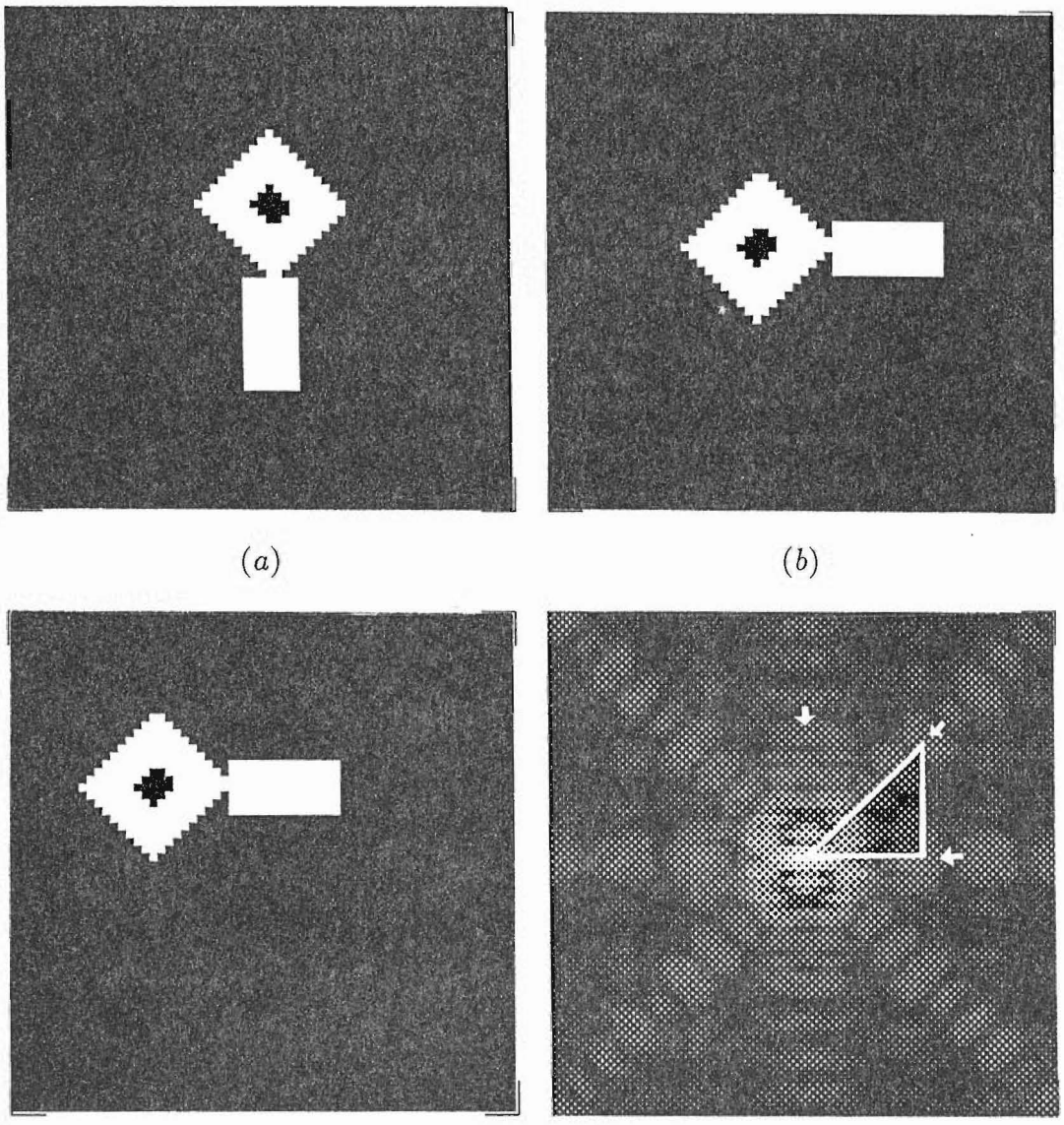
Two computer-generated images invoked in this chapter to illustrate the performance of each of the correlation techniques are introduced in this section.

The first image invoked is shown in Fig. 3.2 which, as an example for illustrating bispectral angular correlation, is described in §3.3.

The second image is shown in Fig. 5.1. It is set in a 64×64 pixel frame (as are all the other images presented in this section). The original image, which is a computer-generated pattern, and a rotated version of it are shown in Fig. 5.1(a) and Fig. 5.1(b) respectively. Fig. 5.1(c) shows a shifted and rotated version. The magnitudes of the spectra of the images shown in, respectively, Fig. 5.1(a) and either Fig. 5.1(b) or Fig. 5.1(c) are displayed in Fig. 5.1(d) and Fig. 5.1(e).

The two objects shown, respectively, in Fig. 3.2(a) and Fig. 5.1(a) are, from now on, referred to as Object-A and Object-B.

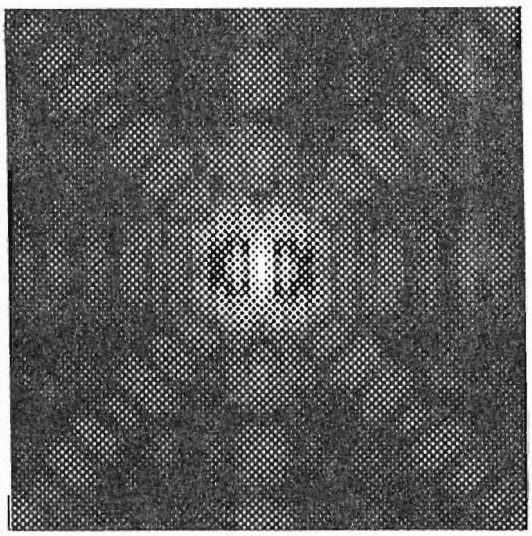
For convenience, two images, from which the relative orientation (rotation) of the



(a) (b)

(c) (d)

Figure continues on next page....



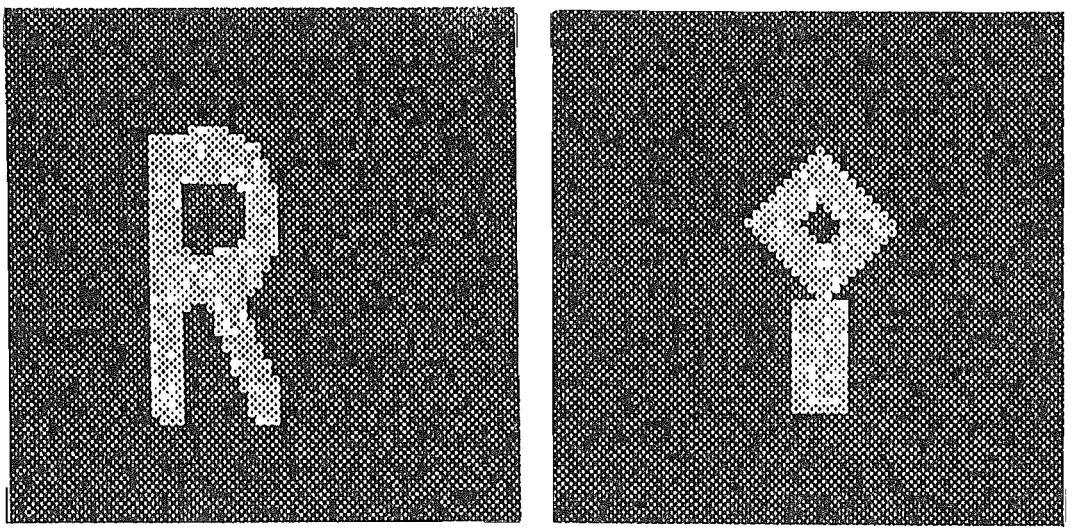
(e)

Figure 5.1. (a) the computer-generated binary image which is in a 64 × 64 pixel frame; (b) the rotated (by 90°) version of (a); (c) the shifted (8 pixels upward and 8 pixels rightward, i.e. the shifts are $a = b = s$, where s represents a constant shift of 8 pixels) and then rotated (by 90°) version of (a); (d) the magnitude of the spectrum of (a); (e) the magnitude of the spectrum of (b) or (c) (spectral magnitudes are shift-invariant). (a) and (b) or (a) and (c) are typical image-pairs. The particular triangle is highlighted in (d) to indicate the part of that bispectrum that is used for rotation estimation.

object is to be determined by the angular correlations, are referred to as members of an image-pair, and are represented respectively by $f(\mathbf{x})$ and $g(\mathbf{x})$. For example, Figs. 3.2(a) and (b) or Figs. 5.1(a) and (c) are two such (uncontaminated) image-pairs. Also note that a and b are the two Cartesian components of the object's shift vector $\mathbf{s}$ between two members of an image-pair, as defined in (1.2). It is seen from Fig. 3.2 and Fig. 5.1 that the shifts for both Object-A and Object B are $a = b = s$, which is 8 pixels as defined in the captions to both Fig. 3.2 and Fig. 5.1.

In order to study the performance of angular correlations in the presence of appreciable contamination, the two image-pairs shown in Fig. 3.2(a), (b) and Fig. 5.1(a), (c) are contaminated with the types of computer-generated backgrounds described in the next three paragraphs.

In the first case, the background image is constructed by choosing each pixel value independently from a uniform pseudo-random distribution, ranging from zero to a positive number. While the average of the background for Object A is set to twice the average of the pixel values of the original image, that of the background for Object B is set to three times the average of the pixel values the original image. The average is here defined as the sum of all $64 \times 64 = 4096$ pixel values divided by 4096. A contaminated image is formed by adding a background to the original image of the object. The two members of an image-pair are contaminated with statistically independent random numbers (although having the same statistics). Shown in Fig. 5.2(a) and (b)



(a) (b)

Figure 5.2. Typical backgrounds (representative of both the first and the second types of contaminations) superimposed upon (a) image of Object-A. (b) image of Object-B.

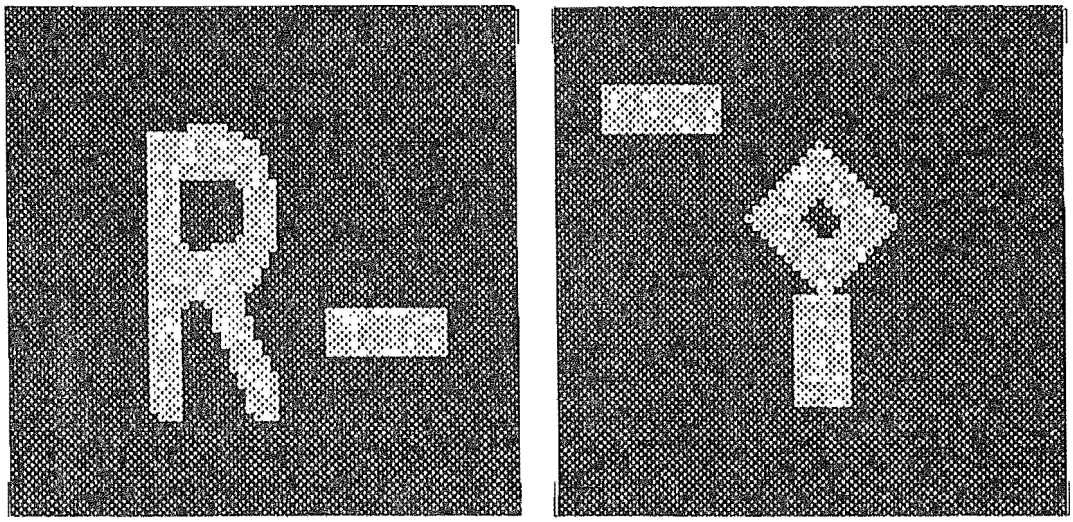
are typical contaminated images of this kind for, respectively, Object A and Object B, where the objects shown are neither rotated nor shifted.

In the second situation, the contamination is the same as in the first except that the two members of an image-pair are contaminated with identical backgrounds, by which is meant that although the object rotates and shifts the background does not, i.e. it is stationary. Fig. 5.2 shows examples of Object A and Object B contaminated in this fashion.

In the third case, the image-pair is first contaminated in a way similar to that for the first case, then the image of an extra rectangular object is superimposed upon each contaminated image. The position and orientation of the extra object are invariant (i.e. the extra object remains stationary while the original object moves). Typical contaminated images are shown in Fig. 5.3(a) and (b) for, respectively, Object A and Object B.

5.2 Rotation Estimation

As mentioned in Chapter 2, spectral values at cylindrical polar coordinates ρ and ϕ are needed for angular correlations. $F(\rho;\phi)$ is usually interpolated from samples obtained at nodes of a rectangular mesh which are most conveniently labelled with Cartesian coordinates. Bispectral values at polar coordinates ρ, ϕ, τ and R can be computed from (3.10) when $F(\rho;\phi)$, $F(\tau;\omega)$ and $F(R;\Theta)$ are all interpolated. The bilinear interpolation scheme introduced in §2.3.2 is invoked for calculating spectral



(a) (b)

Figure 5.3. Typical background of the third type of contamination superimposed upon (a) image of Object-A. (b) image of Object-B.

values at required polar coordinates. An angular resolution of $180/\pi$ is employed here.

Since the two angular correlations described in, respectively, §2.3.6 and §3.3 make use of only partial spectral or bispectral information, it is very important that only the appropriate parts of them are computed. While §5.2.1 discusses the the ordinary spectral method, §5.2.2 details the bispectral technique. Angular correlations performed in the absence of contamination are illustrated in Fig. 5.4 which corresponds to Fig. 5.1.

5.2.1 Ordinary Spectral Rotation Estimations

Recalling the discussion in §2.3.6, in which only object rotation is envisaged, (2.45) can be invoked, i.e. both the magnitude and the phase information can be employed for estimating the rotation. It is pointless computing $\Gamma(F_\rho, G_\rho, \phi)$ for values of ρ for which $|F(\rho;\phi)|$ is small for all ϕ. Furthermore, it is usually a waste of computational effort to examine values of ρ spanning the central lobe of $F(\rho;\phi)$, because both the magnitude and phase of the latter tend to vary slowly with ϕ within this lobe. It is, in fact, only worthwhile to make calculations for values of ρ corresponding to the peaks of fringes (or sidelobes) of $F(\rho;\phi)$ outside the central lobe. For illustration, the top curve in Fig. 5.4 shows $\Gamma(F_\mu, G_\mu, \phi)$ where μ is the value of ρ corresponding to the peak of $|F(\rho;\phi)|$ marked with a horizontal arrow in Fig. 5.1(d).

The top curve in Fig. 5.4 applies to a purely rotating object (corresponding to

Fig. 5.1(a) and (b)), for which $a = b = 0$. The rotation of $g(\mathbf{x})$ with respect to $f(\mathbf{x})$ is 90° ($\psi = \pi/2$).

The second curve from top in Fig. 5.4 applies to an object (corresponding to Fig. 5.1(a) and (c)) which both rotates ($\psi = \pi/2$) and shifts, with a and b equal to one eighth of the side of the square frame enclosing Fig. 5.1(a), i.e. 8 pixels. Whereas the top curve clearly peaks at $\phi = \pi/2$, thereby correctly identifying the object's rotation, the rotation estimate provided by the second curve from the top is in error by close to $7\pi/4$ because of the presence of the term $\exp[-i2\pi[au + bv)]$ on RHS of (2.20).

As discussed in §2.3.6, the effect of the object's shift on the angular correlation can be removed by discarding the spectral phases and computing $\Gamma(|F_\mu|, |G_\mu|, \phi)$, which is shown in Fig. 5.4 as the third curve from the top. The major peak of the latter curve, while occurring at the same value of ϕ as for the top curve, is considerably broader. Even more importantly, the sidelobes of $\Gamma(|F_\mu|, |G_\mu|, \phi)$ are much higher than those of $\Gamma(F_\mu, G_\mu, \phi)$. In the presence of variable background, therefore, provided the effects of shifts can be nullified, the object's rotation is likely to be less accurately estimated by $\Gamma(|F_\mu|, |G_\mu|, \phi)$ than by an estimator which retains Fourier phase information. Just such an estimator, the bispectral angular correlation, is introduced in the next subsection.

5.2.2 Bispectral Rotation Estimation

Since the bispectra are shift-invariant, their phases can be usefully retained in the integrand of (3.16). The requirements for angularly correlating two spectral magnitudes described in the first paragraph of the last subsection also apply to bispectral angular correlations. Reference to (3.7) and (3.16) indicates that the bispectral correlation $\Gamma(F_{\rho;\tau;R}, G_{\rho;\tau;R}, \phi)$ can only be a sensitive estimator of the object's rotation if $|F(\mathbf{u})|$, $|F(\boldsymbol{\alpha})|$ and $|F(\mathbf{u} + \boldsymbol{\alpha})|$ are simultaneously of appreciable amplitude for at least one pair of values of $\mathbf{u}$ and $\boldsymbol{\alpha}$. When $\phi = 0$, ρ is also set to the value μ, as introduced in the last subsection and mentioned in the caption to Fig. 5.1. τ and R are taken to be, respectively, ν and Υ such that $\mu = \pi/2$ where $(\mu, \pi/2)$ is the point in Fig. 5.1(d) identified by a vertical arrow. The point in Fig. 5.1(d) indicated by a diagonal arrow corresponds to $(\mathbf{u} + \boldsymbol{\alpha})$ when the points marked by horizontal and vertical arrows correspond to $\mathbf{u}$ and $\boldsymbol{\alpha}$ respectively. These three positional vectors, i.e. $\mathbf{u}$, $\boldsymbol{\alpha}$ and $(\mathbf{u}+\boldsymbol{\alpha})$ relate to each other in the form of the congruent triangle in the spectral space, as illustrated in Fig. 5.1(d). The bottom curve in Fig. 5.4 shows the bispectral angular correlation $\Gamma(F_{\mu,\nu,\Upsilon}, G_{\mu,\nu,\Upsilon}; \phi)$ (corresponding to Fig. 5.1(a) and (c)) for $\psi = \pi/2$.

The major peak of the bottom curve in Fig. 5.4 is appreciably narrower than the major peaks of the other three curves, implying that the bispectral estimator is capable of finer resolution than either $\Gamma(|F_\mu|, |G_\mu|; \phi)$ or $\Gamma(F_\mu, G_\mu; \phi)$. Furthermore, this resolution improvement persists when the object not only rotates but also shifts.

Performance of the ordinary spectral magnitude and bispectral estimators in the

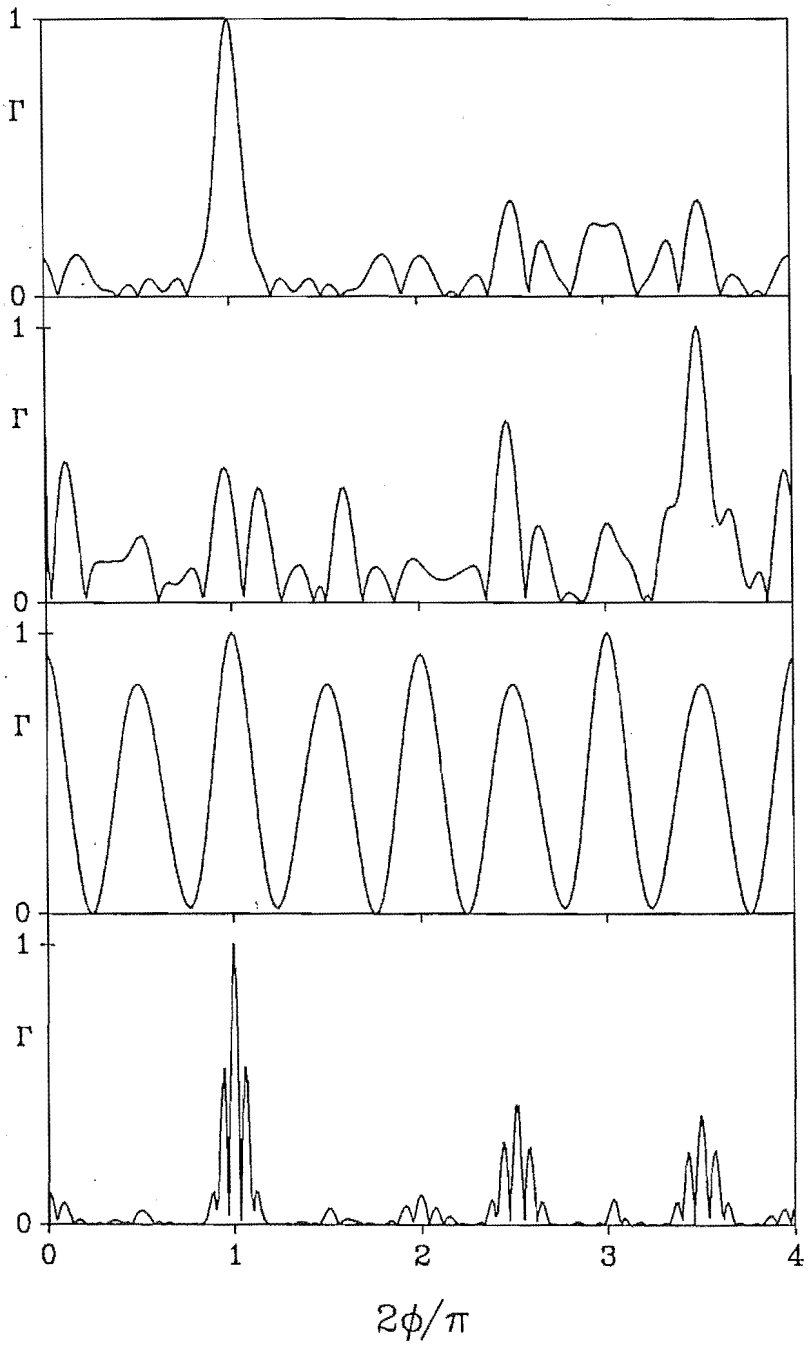


Figure 5.4. Normalized (to maximum values of unity) angular correlations Γ for ordinary-spectral and bispectral estimators ($\psi = \pi/2$, μ, ν and Υ as indicated in Fig. 5.1(d) and outlined in §5.2.1):

Top curve :	$\Gamma(F_\mu, G_\mu; \phi)$ with $a = b = 0$
Second from top :	$\Gamma(F_\mu, G_\mu; \phi)$ with $a = b = s$
Third from top :	$\Gamma(\lvert F_\mu \rvert, \lvert G_\mu \rvert; \phi)$ with $a = b = s$
Bottom curve :	$\Gamma(F_{\mu,\nu,\Upsilon}, G_{\mu,\nu,\Upsilon}; \phi)$ with $a = b = s$

Note that s equals one eighth of the side of the square enclosing $f(\mathbf{x})$ as indicated in the caption to Fig. 5.1.

presence of appreciable backgrounds are compared in the next section.

5.3 Rotation/Shift Estimation

In this section, comparisons are presented for two algorithms which first estimate the rotation, then compensate for the estimated rotation, and finally estimate the shift. The phase correlation procedure described in §2.3.4 is incorporated into both algorithms. Conventional cross correlation has also been invoked, and produced almost identical results. So as to make the comparison of the two algorithms realistic, the three types of contaminated images described in §5.1 are employed.

The computation of the spectral magnitudes and the bispectra for the contaminated images is outlined as follows. For Object-B (Fig. 5.1(a) and (c)), as for the situation of contamination-free discussed in §5.2, ρ, τ and R are chosen to be, respectively, μ, ν and Υ, as indicated in Fig. 5.1(d) for the three types of contaminated images illustrated in Fig. 5.2(b) and Fig. 5.3(b) respectively. For Object-A (Fig. 3.2(a) and (b)), ρ, τ and R are taken to be, respectively, κ, λ and χ as indicated in Fig. 3.2(c), where, the three points indicated by, respectively, the horizontal, vertical and diagonal arrows correspond to $\mathbf{u}$, $\boldsymbol{\alpha}$ and $(\mathbf{u}+\boldsymbol{\alpha})$. The corresponding three types of contaminated images are illustrated in Fig. 5.2(a) and Fig. 5.3(a) respectively.

For each algorithm, after an estimate $\hat{\psi}$ for the rotation angle has been generated, the rotation compensated Fourier transform $\tilde{G}(\rho;\phi)$ is defined by

$$\tilde{G}(\rho;\phi) = G(\rho;\phi+\psi) = \tilde{G}(\mathbf{u}) \tag{5.1}$$

On defining $\tilde{g}(\mathbf{x}) \longleftrightarrow \tilde{G}(\mathbf{u})$, it seen that $\tilde{g}(\mathbf{x})$ is a version of $g(\mathbf{x})$ compensated for the rotation. In order to generate an estimate $\tilde{\mathbf{s}}$ of the shift, one computes the phase correlation defined by

$$c(\mathbf{x}) \longleftrightarrow W(\mathbf{u}) \exp j\left\{\text{Phase}\{\tilde{G}(\mathbf{u})\} - \text{Phase}\{F(\mathbf{u})\}\right\} \tag{5.2}$$

where $W(\mathbf{u})$ is the Blackman window as described in §4.2.1. The estimated shift $\tilde{\mathbf{s}}$, having Cartesian components $\tilde{a}$ and $\tilde{b}$, is then defined to be that $\mathbf{x}$ for which $|c(\mathbf{x})|$ is greatest.

5.3.1 Ordinary-Spectral Magnitude Algorithm

This algorithm starts by computing the magnitudes $|F_\rho(\phi)|$ and $|G_\rho(\phi)|$ of, respectively, the spectra of $f(\mathbf{x})$ and $g(\mathbf{x})$. Then the correlation $\Gamma(|F_\rho|,|G_\rho|,\phi)$, for the value of ρ chosen in the way described in the second introductory paragraph of §5.3, is evaluated. The value of $\tilde{\psi}$ is taken to be that ϕ for which the $\Gamma(|F_\rho|,|G_\rho|,\phi)$ is greatest. This procedure is repeated for each image-pair being studied.

5.3.2 Bispectral Algorithm

This algorithm starts by computing the bispectra $F_{\rho,\phi,R}(\phi)$ and $G_{\rho,\tau,R}(\phi)$ of, respectively, $f(\mathbf{x})$ and $g(\mathbf{x})$. Then the correlation $\Gamma(F_{\rho,\tau,R}, G_{\rho,\tau,R}, \phi)$, for the values of ρ, τ and R chosen in the way described in the second introductory paragraph of §5.3, is evaluated. The value of $\tilde{\psi}$ is taken to be that ϕ for which the bispectral correlation $\Gamma(F_{\rho,\tau,R}, G_{\rho,\tau,R}, \phi)$ is greatest. This procedure is repeated for each image-pair being studied.

5.3.3 Comparison of Algorithms' Performance

The results of several comparisons of the above two algorithms are presented here. The data for each comparison consists of an image-pair, comprising contaminated versions of $f(\mathbf{x})$ and $g(\mathbf{x})$. For each image-pair that is a contaminated version of either Fig. 3.2(a) and (d) or Fig. 5.1(a) and (c), the object rotation is $\psi = \pi/2$, and the shifts are $a = b = s$ (with s being 8 pixels), as defined in the captions to Fig. 3.2 and Fig. 5.1.

There are three sets of comparisons corresponding to the three types of contaminations described in §5.1 for each of the two objects, Object-A and Object-B, respectively. For each set of estimates, a pair of typical angular correlations are plotted corresponding to the spectral and the bispectral methods respectively. The arrangement for presenting the results is:

Object-A Fig. 5.5 and the first row of Table 5.1 correspond to the first set of comparisons. Fig. 5.6 and the second row of Table 5.1 correspond to the second set of comparisons. Fig. 5.7 and third row of Table 5.1 correspond the third set of comparisons.

Object-B Fig. 5.8 and the fourth row of Table 5.2 correspond to the first set of comparisons. Fig. 5.9 and the fifth row of Table 5.2 correspond to the second set of comparisons. Fig. 5.10 and sixth row of Table 5.2 correspond the third set of comparisons.

Each set comprises 20 contaminated image-pairs differing from each other only in their backgrounds, which are statistically independent (although having the same statistics). It is found that the number of image-pairs per set is sufficient to give a useful idea of how well the algorithms perform in the presence of appreciable contamination. Because the extra object in both Fig. 5.3(a) and Fig. 5.3(b) is smaller then the original object, the latter's rotation and shift are expected to be estimated by the algorithms, with the former consitituting additional contamination.

Figs. 5.5, 5.6 and 5.7 show $\Gamma(|F_\lambda|, |G_\lambda|; \phi)$ and $\Gamma(F_{\lambda,\kappa,\chi}, G_{\lambda,\kappa,\chi}; \phi)$ for typical image-pairs belonging to each of the three sets for Object-A. For this example, the spectral magnitude and bispectral algorithms perform almost equally well in the presence of appreciable background, as is confirmed by the results collected in Table 5.1, which lists the number of image-pairs whose rotation and shifts are estimated with very

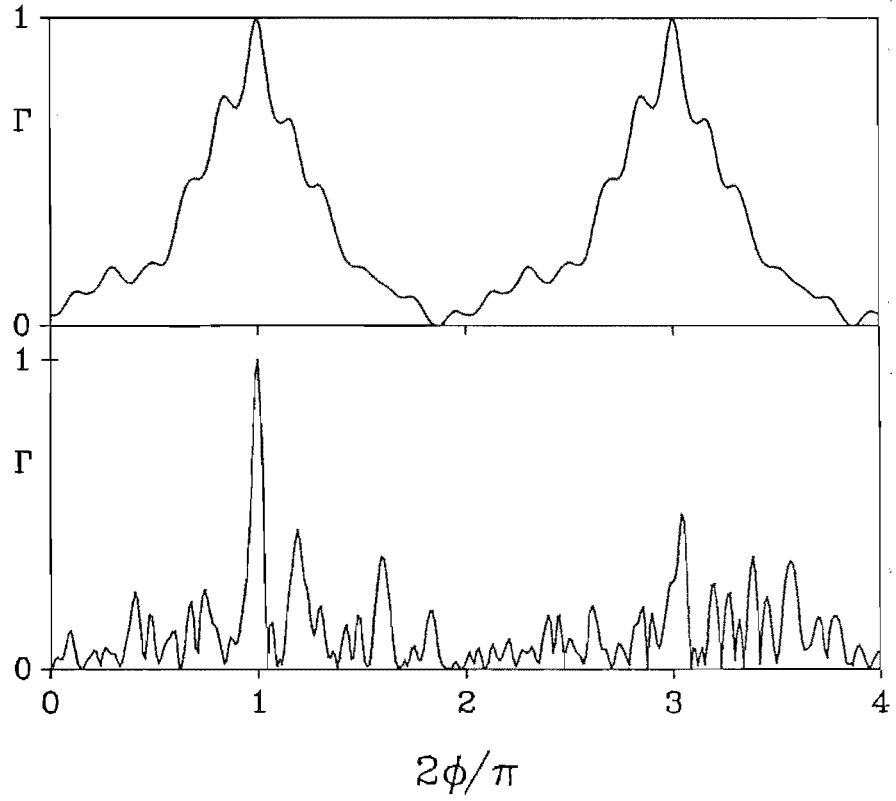


Figure 5.5. Typical normalized (to maximum value of unity) angular correlations Γ for the first set of image-pairs corresponding to Object-A ($\psi = \pi/2, \kappa, \lambda$ and χ as indicated in Fig. 3.2(c) and specified in the second paragraph of §5.3):
Upper: $\Gamma(|F_\kappa|, |G_\kappa|; \phi)$ with $a = b = s$
Lower: $\Gamma(F_{\kappa,\lambda,\chi}, G_{\kappa,\lambda,\chi}; \phi)$ with $a = b = s$.

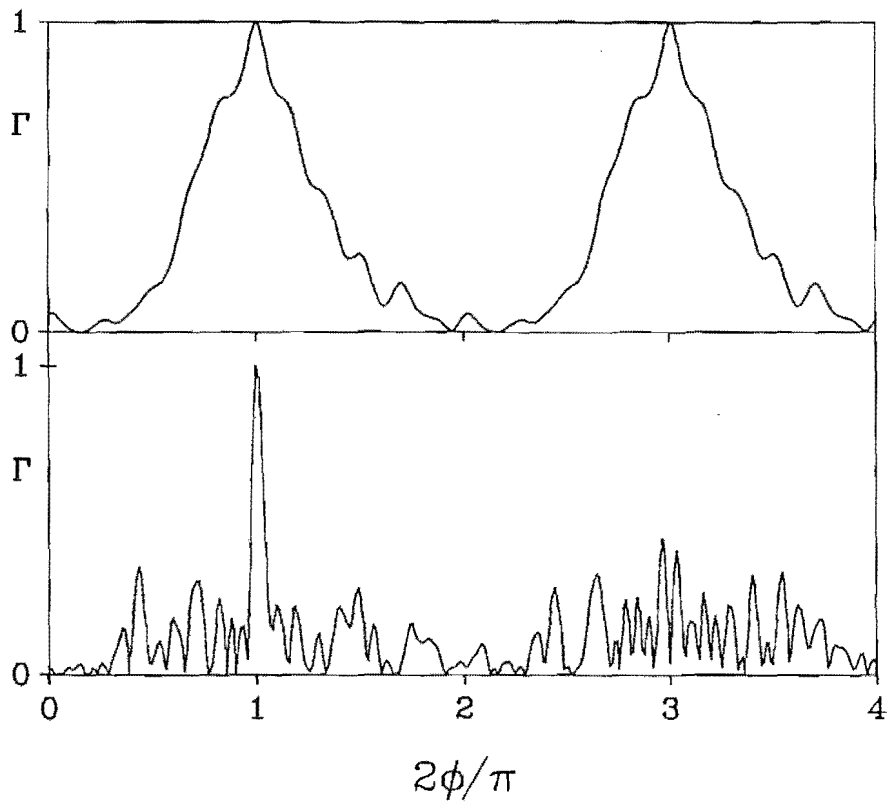


Figure 5.6. Typical normalized (to maximum value of unity) angular correlations Γ for the second set of image-pairs corresponding to Object-A ($\psi = \pi/2, \kappa, \lambda$ and χ as indicated in Fig. 3.2(c) and specified in the second paragraph of §5.3):
Upper: $\Gamma(|F_\kappa|, |G_\kappa|; \phi)$ with $a = b = s$
Lower: $\Gamma(F_{\kappa,\lambda,\chi}, G_{\kappa,\lambda,\chi}; \phi)$ with $a = b = s$.

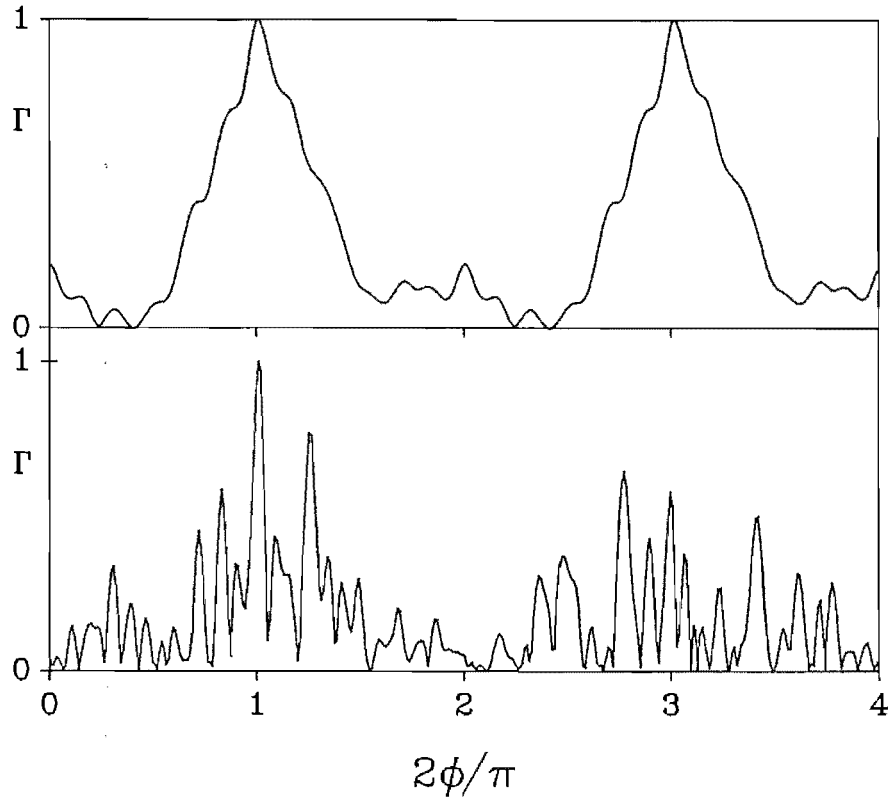


Figure 5.7. Typical normalized (to maximum value of unity) angular correlations Γ for the third set of image-pairs corresponding to Object-A ($\psi = \pi/2, \kappa, \lambda$ and χ as indicated in Fig. 3.2(c) and specified in the second paragraph of §5.3):
Upper: $\Gamma(|F_\kappa|, |G_\kappa|; \phi)$ with $a = b = s$
Lower: $\Gamma(F_{\kappa,\lambda,\chi}, G_{\kappa,\lambda,\chi}; \phi)$ with $a = b = s$.

Algorithm	*Ordinary–Spectral Magnitude*		*Bispectral*	
Set of image–pairs	*Rotation*	*Shift*	*Rotation*	*Shift*
First	19	20	20	20
Second	18	20	19	20
Third	16	20	16	20

Table 5.1. Numbers of image-pairs in each of the three sets (corresponding to Object-A) for which the estimates of rotation and shifts are within 1° and $0.125s$ (1 pixel), respectively, of the actual rotation and shift, where s is defined in the caption to Fig. 3.2. Note that only 12 rotation estimates are out by as much as twice the error limit. All shifts are correctly estimated since phase correlation is relatively insensitive to changes in rotation angle of up to 4°, as illustrated in Fig. 4.15.

Algorithm	*Ordinary–Spectral Magnitude*		*Bispectral*	
Set of image–pairs	*Rotation*	*Shift*	*Rotation*	*Shift*
First	8	9	19	19
Second	5	6	20	20
Third	0	0	15	15

Table 5.2. Numbers of image-pairs in each of the three sets (corresponding to Object-B) for which the estimates of rotation and shifts are within 2° and $0.125s$ (1 pixel), respectively, of the actual rotation and shift, where s is defined in the caption to Fig. 5.1. Note that the ordinary spectral magnitude algorithm estimates shifts very slightly better than rotations (within the arbitrarily chosen error limits). All but 6 of the estimates not included in this table are in error by more than 40° in rotation and $0.5s$ (4 pixels) in shift. The other 6 estimates are in error by roughly 4° and $0.125s$.

high accuracy by the algorithms. While all shift estimates are within the error limit ($0.125s$) quoted in the caption to Table 5.1, 12 rotation estimates are out by as much as twice the error limit (1°). However, the major peak of $\Gamma(F_{\lambda,\kappa,\chi}, G_{\lambda,\kappa,\chi}; \phi)$ is still much narrower then that of $\Gamma(|F_\lambda|, |G_\lambda|; \phi)$ in each of Figs. 5.5, 5.6 and 5.7.

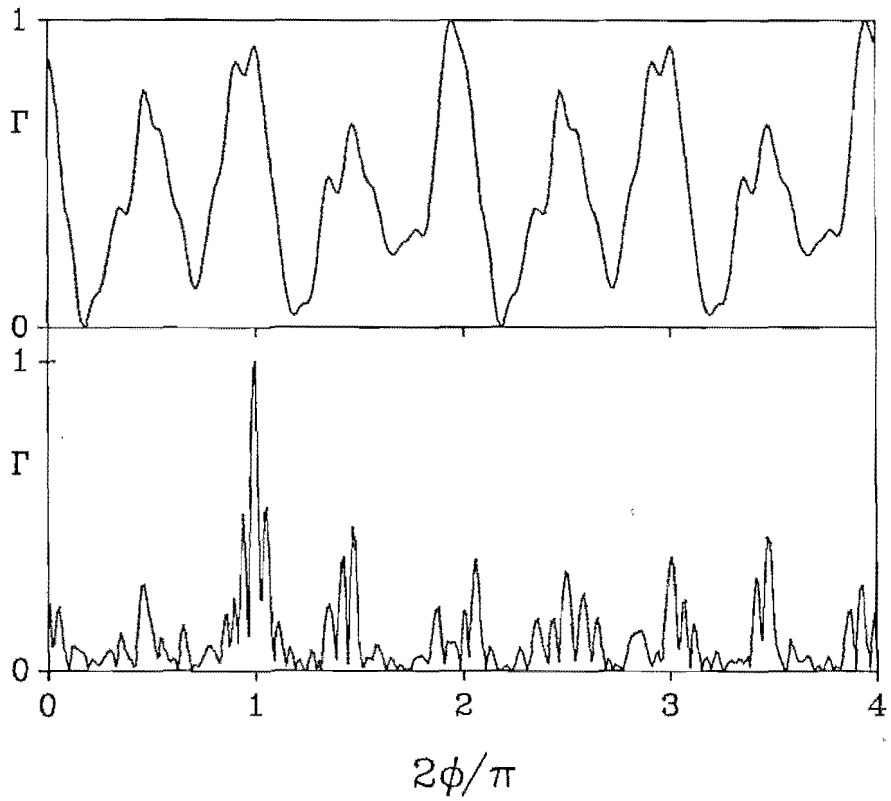


Figure 5.8. Typical normalized (to maximum value of unity) angular correlations Γ for the first set of image-pairs corresponding to Object-B ($\psi = \pi/2, \mu, \nu$ and Υ as indicated in Fig. 5.1(d) and specified in the second paragraph of §5.3):
Upper: $\Gamma(|F_\mu|, |G_\mu|; \phi)$ with $a = b = s$
Lower: $\Gamma(F_{\mu,\nu,\Upsilon}, G_{\mu,\nu,\Upsilon}; \phi)$ with $a = b = s$.

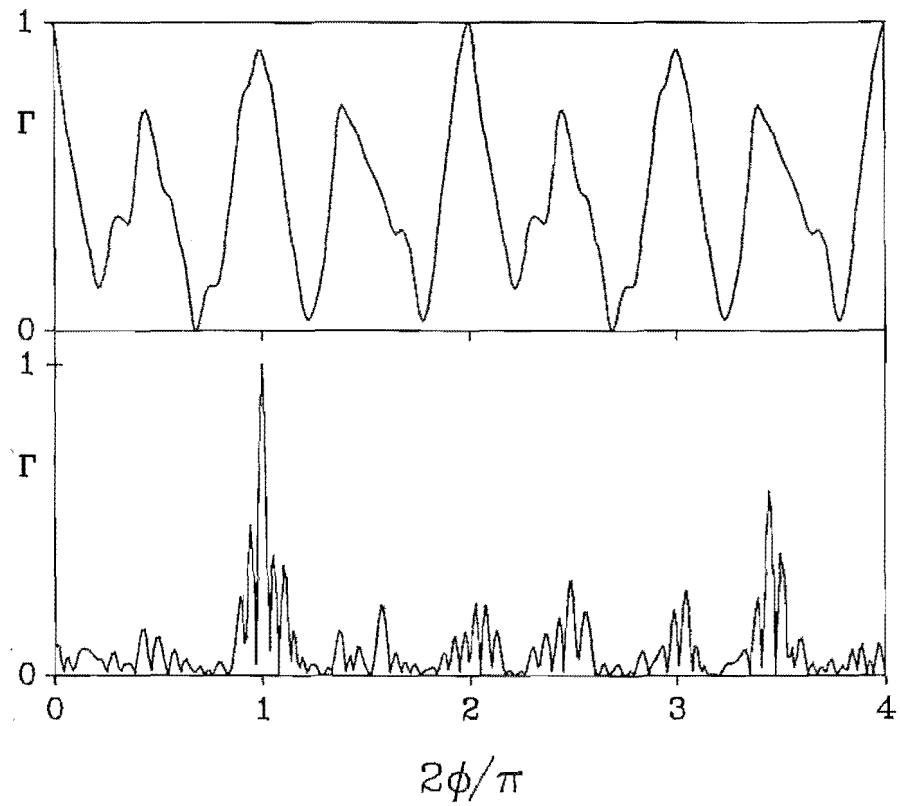


Figure 5.9. Typical normalized (to maximum value of unity) angular correlations Γ for the second set of image-pairs corresponding to Object-B ($\psi = \pi/2, \mu, \nu$ and Υ as indicated in Fig. 5.1(d) and specified in the second paragraph of §5.3):
Upper: $\Gamma(|F_\mu|, |G_\mu|; \phi)$ with $a = b = s$
Lower: $\Gamma(F_{\mu,\nu,\Upsilon}, G_{\mu,\nu,\Upsilon}; \phi)$ with $a = b = s$.

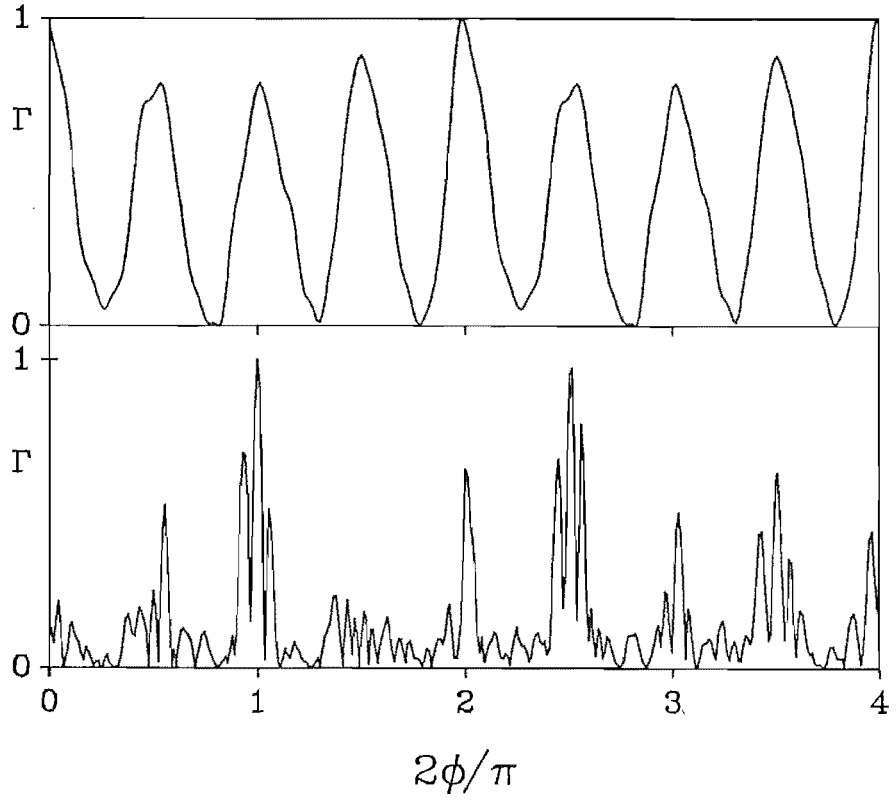


Figure 5.10. Typical normalized (to maximum value of unity) angular correlations Γ for the third set of image-pairs corresponding to Object-B ($\psi = \pi/2, \mu, \nu$ and Υ as indicated in Fig. 5.1(d) and specified in the second paragraph of §5.3):
Upper: $\Gamma(|F_\mu|, |G_\mu|; \phi)$ with $a = b = s$
Lower: $\Gamma(F_{\mu,\nu,\Upsilon}, G_{\mu,\nu,\Upsilon}; \phi)$ with $a = b = s$.

Figs. 5.8, 5.9 and 5.10 show $\Gamma(|F_\mu|, |G_\mu|; \phi)$ and $\Gamma(F_{\mu,\nu,\Upsilon}, G_{\mu,\nu,\Upsilon}; \phi)$ for typical image-pairs belonging to each of the three sets for Object-B. In contrast to the results obtained for Object-A, the marked superiority of the bispectral algorithm in the presence of more significant backgrounds, as suggested by these three figures, is reinforced by the results collected in Table 5.2, which lists the number of image-pairs whose rotations and shifts are estimated with considerable accuracy by the algorithms. It is worthy of emphasis that the great majority of estimates are either highly accurate or wildly inaccurate. Of the 104 estimates which exceed the error limits quoted in the caption to Table 5.2, 98 are so erroneous as to be worthless from any practical point of view. The other 6 estimates are only out by approximately twice the quoted error limits. The ordinary-spectral magnitude algorithm is seen to be inadequate for over 50% of image-pairs, whereas only for the third set of image-pairs does the bispectral algorithm perform less then very well. The improved shift estimates provided by the bispectral algorithm are due to it compensating more faithfully for rotation (implying that the PC procedure described in §2.3.4 has better data to operate upon). Whereas the major peaks of $\Gamma(|F_\mu|, |G_\mu|; \phi)$ in each of Figs. 5.8, 5.9 and 5.10 occurs near to $\phi = \pi$, all three versions of $\Gamma(F_{\mu,\nu,\Upsilon}, G_{\mu,\nu,\Upsilon}; \phi)$ peak close to the correct value of $\pi/2$. While it should be noted that the secondary peak in the lower curve of Fig. 5.10 is nearly as high as the main peak, it is worth emphasising that the image-pair giving rise to Fig. 5.10 is doubly contaminated, with a stationary extra object as well as variable backgrounds.

5.4 Discussion

It is seen that the results presented in §5.3.3 for the two examples are very different from each other. The reason is, as intimated in §5.1, that the level of contamination is effectively lower for the first example (corresponding to Object-A) than for the second (corresponding to Object-B). Therefore, it seems that both of these algorithms perform well for contamination below a certain level. When the contamination exceeds this level, the results suggest that the algorithm based on the Erlangen-bispectrum is much more robust. This applies as much to the resolution as to the stability of the estimates provided by the algorithm. Despite the high background levels present in all image-pairs of the three sets, for each of the two examples discussed in §5.1, the widths of the major peaks of $\Gamma(F_{\lambda,\kappa,\chi}, G_{\lambda,\kappa,\chi}; \phi)$ in Figs. 5.5, 5.6 and 5.7 as well as those of $\Gamma(F_{\mu,\nu,\Upsilon}, G_{\mu,\nu,\Upsilon}; \phi)$ in Figs. 5.8, 5.9 and 5.10 appear to be as narrow as the major peaks of the bottom curves in both Fig. 3.3 and Fig. 5.4. The major reason for this is that the spectral phase is retained in the bispectral approach.

It is worth emphasising the importance of the spectral phase. Referring to (2.44), it is seen that the spectral magnitude angular correlation operates upon two real and positive functions. When implementing the correlation digitally, any correlation value is a sum of many products of two real and positive numbers. Real and positive numbers always sum additively, of course. On the other hand, the bispectral algorithm operates on a pair of complex functions, as is clear from (3.16). Now consider the outcome of correlating two complex functions which are conjugated and shifted versions of each other (the situation that is envisaged here). Firstly, when the two

functions are exactly matched, the phases of the two functions in the integrand of the correlation integral cancel each other completely, because the two functions are defined to be conjugates of each other. Hence the correlation necessarily reaches a maximum when the functions are exactly matched (equivalent to constructive interference). When the two functions are not closely matched, the phase of the integrand of the correlation is no longer identically zero. It tends, in fact, to vary considerably through the range of integration, leading to cancellation of parts of the integrand with other parts (equivalent to destructive interference). Therefore, in general, bispectral angular correlations tend to be sharply peaked for angles corresponding to the rotations of objects. From a statistical point of view, any two real and positive functions are significantly correlated, while any two complex functions are unlikely to be significantly correlated. Therefore, the angular correlation technique can usually be expected to be more effective when employing the bispectrum rather than the spectral magnitude.

It is also seen, from the results obtained for the second example (see Table 5.2), that the great majority of estimates are either highly accurate or wildly inaccurate. This may be explained as follows. When the contamination exceeds a certain level, there is a significant chance that the major peak moves to a position where, in the absence of contamination, a peak of significant height exists. This becomes clear when one compares the bottom curve of Fig. 5.4 with that of Fig. 5.10, which illustrates how highly inaccurate estimates may arise. However, one may sometimes be able to improve the estimation procedure in the following manner. If the object's rotation ψ is known *a priori* to be within a particular angular interval, only correlation values within this interval should be examined when estimating the rotation. This increases the reliability of estimating rotation angles with the aid of these algorithms. It also reduces computation time since one only needs to compute correlation values for rotation angles within the given interval.

Computational experiments described in this chapter were conducted on a VAX 3500 computer. It takes approximately 8 seconds to complete the procedure for processing an image-pair of 64×64 pixels by the bispectral algorithm, which includes estimating, respectively, rotation by bispectral angular correlation and shift by phase correlation. The time taken for the spectral magnitude algorithm is only slightly less than for the bispectral algorithm. However, if one estimates shift and rotation by methods other than those based on shift-invariant mapping (e.g. by matching with the conventional cross-correlation discussed in §1.2.2), the computational cost is very high for processing an image-pair. This is because a full-frame interpolation and a cross correlation operation are necessary for every possible angle. To obtain the same angular resolution invoked in this chapter, 360 cross correlations need to be computed. The time taken for each cross correlation operation (including a full-frame bilinear interpolation) is about 16 seconds on the same computer. This time can be reduced by invoking the feature based matching techniques on feature extracted images as described in §1.2.3. However, the overall computational requirements remain large, because of the large number of possible rotation angles. Therefore, compared to the spatial matching algorithm, the bispectral or the spectral methods presented in this chapter reduce the computational requirements by a factor of about 100.

Possible applications of the angular correlation techniques described in this chapter to real-world images are discussed in Chapter 8.

Chapter 6

Digital Coronary Angiography

This chapter discusses several aspects of digital coronary angiography, one of the contemporary methods of examining and assessing the function and disease of coronary arteries. As an introduction, cardiac function, diagnosis of cardiac disease and the development of digital coronary angiography are briefly described in §6.1. Image formation and quality (in terms of resolution and noise characteristics) are discussed in §6.2. Finally, §6.3 discusses techniques of digital angiographic processing and their application to diagnosing coronary disease. The discussion in §6.3 emphasises the artery motion problem that has to be solved successfully if these techniques are to help with the diagnosis of coronary disease.

6.1 Introduction

6.1.1 Cardiac Function and Abnormalities

The heart is a muscular contractile organ serving as the pump for the circulatory system. It is a four-chambered pump consisting of two (left and right) ventricles, which are the main pumping chambers, and two (left and right) atria, which are the reservoirs that collect blood returning to the heart and supply it to the ventricles (see Fig. 6.1). The walls of these chambers are constructed mainly of a specialized type of muscle called the *myocardium*. The myocardium is mainly made of *cardiac cells* which contract and rest alternatively, so that the heart can pump. This process is governed by the electrical conduction mechanism of the heart [Bones, 1981; Hurst and Logue, 1978].

Blood (and thereby oxygen) is supplied to the myocardium by the coronary arteries [Hurst and Logue, 1978]. The left and right coronary arteries spread over the outer surface (epicardium) of the heart (see Fig. 6.2). The left coronary artery splits into two branches (the left anterior descending and circumflex arteries) shortly after its origin at the aorta (Fig. 6.2).

Cardiac abnormalities can be broadly separated into two categories: congenital (inherited) and acquired (due to aging or disease processes) [Hurst and Logue, 1978].

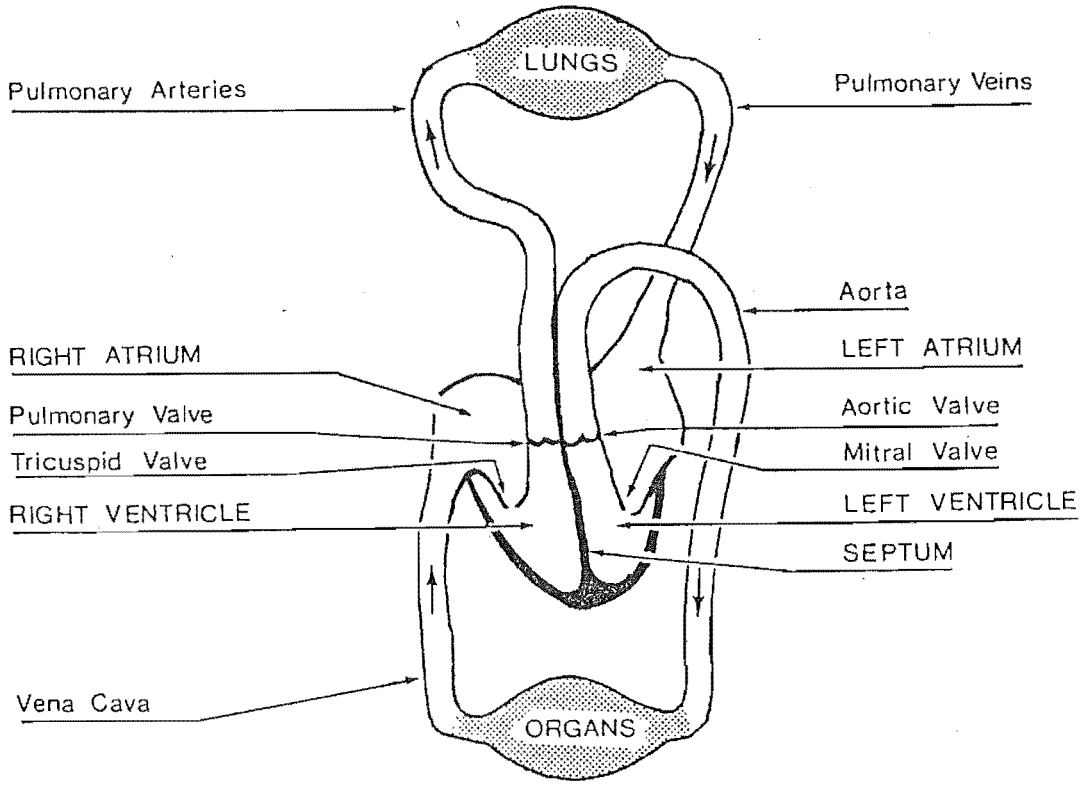
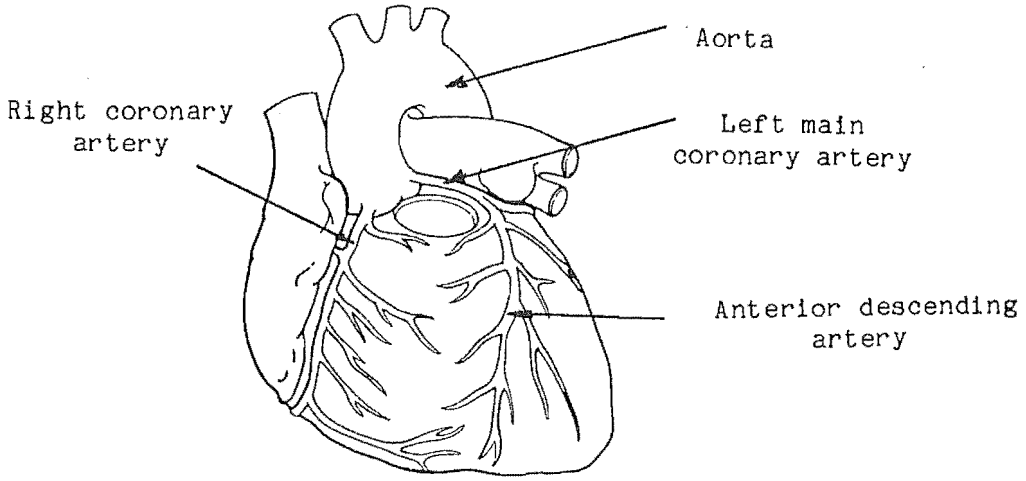


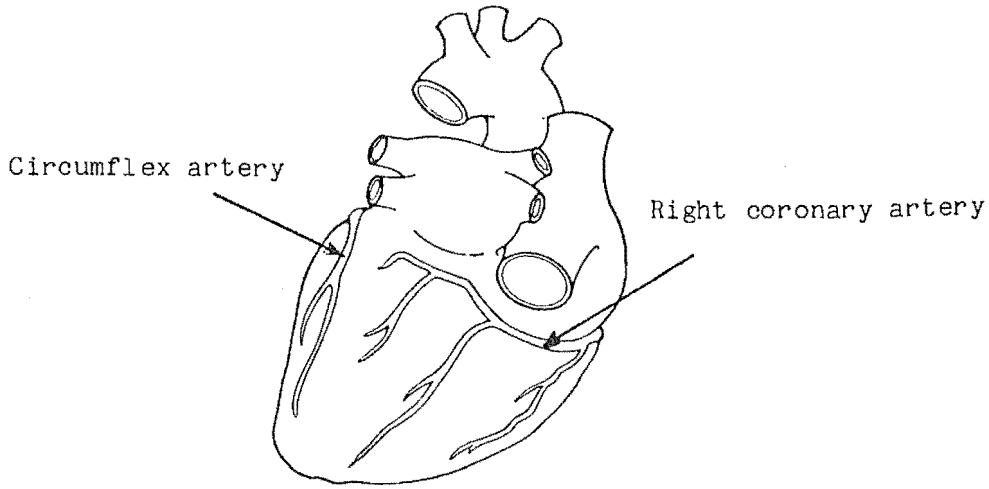
Figure 6.1. The function of the human heart. Blood enters the atria and is pumped out of the ventricles.

Congenital defects usually arise because a malformation occurs at some stage during the development of the fetal heart. Generally, the earlier in the fetal development that this occurs, the more serious is the defect [Hurst and Logue, 1978]. Much more commonly encountered is the problem of acquired heart disease. The most common disease in this category is that associated with the heart's own circulation, i.e. the coronary disease which is currently the single largest cause of death in the Western world [McAinsh, 1986].

Disease of the coronary arteries [Hurst and Logue, 1978] causes the common condition of angina pectoris, a crushing pain in the chest brought on typically by exercise. This pain is a result of the tissue supplied by the vessel being starved of blood. Coronary disease usually takes the form of atheroma, in which fatty, granular deposits develop in the arterial wall, narrowing the artery and causing hardening. Complete occlusion of the vessel (blocking of blood flow) causes an event known as a *myocardial infarction* or, more commonly, a *heart attack*, which can also be caused by a clot which forms in the coronary arteries and deprives the cardiac cells of blood (*coronary thrombosis*). The tissue subsequently dies (necrosis) and is replaced by scar tissue (fibrosis) which cannot contract, hence impairing the pumping performance of the ventricles. Fig. 6.3 illustrates this process. Fig. 6.4 and Fig. 6.5 show coronary angiograms of, respectively, normal and diseased arteries.



(a) front view



(b) back view

Figure 6.2. The structure of coronary arteries: (a) front and (b) back views of the coronary arterial tree on the surface of the heart. The coronary arteries originate from the aorta and penetrate the heart to supply the myocardium with blood.

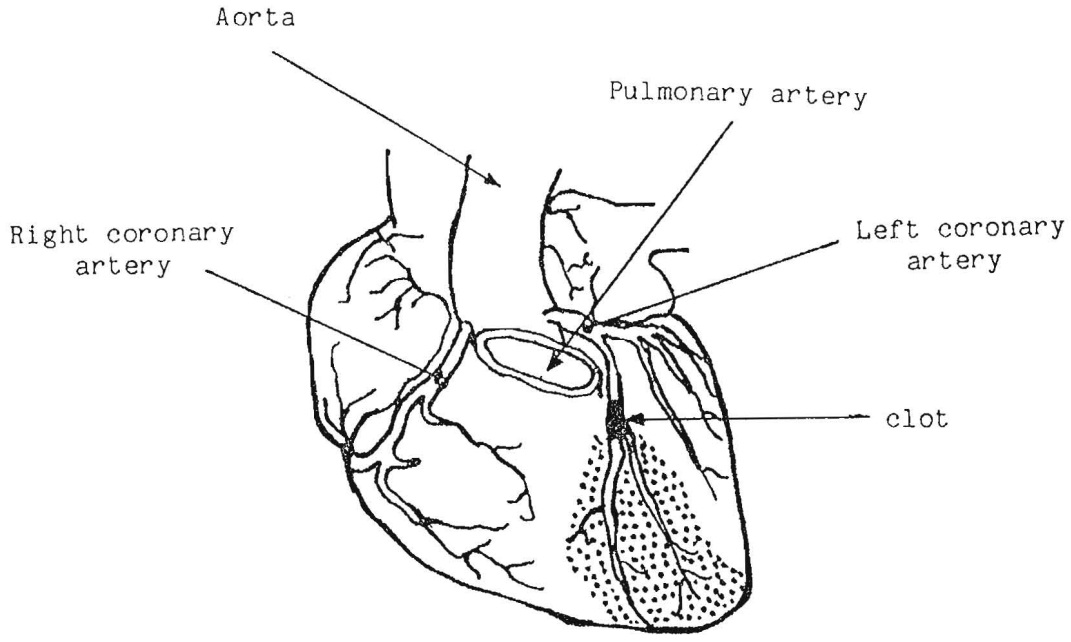


Figure 6.3. Illustration of coronary disease process. The clot or fatty, granular deposits in the coronary artery blocks the blood flow and causes heart attack. Note that the dotted region illustrates the area which is suffering from the lack of blood supply.

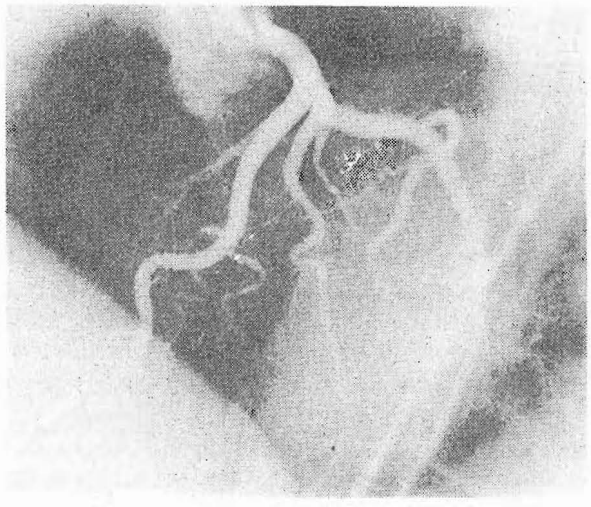


Figure 6.4. Coronary angiogram of a normal coronary artery. There are no narrowings on the coronary arterial tree.

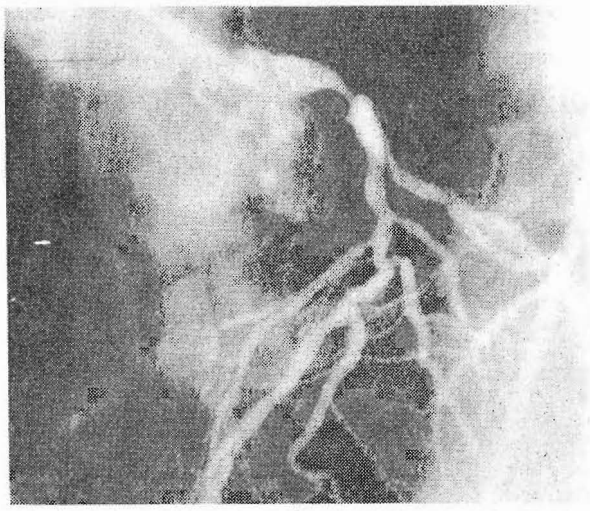


Figure 6.5. Coronary angiogram of an abnormal coronary artery. Narrowings occur at several places of the coronary arterial tree.

6.1.2 Diagnosis and Assessment of Cardiac Disease

The most commonly invoked technique for diagnosing cardiac disease is *angiography*, which relies on diagnostic X-ray equipment. The main components of a diagnostic X-ray system are the X-ray tube, image intensifier, and cine/video camera for recording images onto film/tape, as indicated in Fig. 6.6. In conventional angiography, the X-ray contrast medium (dye) is selectively injected into a certain part of the human circulatory system for visualizing its structure. Images so obtained are subsequently called contrast-enhanced images. Cardiac catheterization, in which a very narrow, flexible and hollow tube is inserted into a region of the heart, is routinely invoked for injecting dye into a heart chamber or even a branch of the coronary arteries through a peripheral artery or vein in order to establish the existence and exact location of a diseased or defected area [Maddison, 1977]. All assessment during catheterization is instrumentation-based and highly invasive since access to the central circulation (the heart) is an unavoidable prerequisite. There is a finite risk associated with performing catheterization (one patient in about 1000 dies). The improvement in diagnosis resulting from the use of such procedures must always be balanced against the physical damage and mental trauma they produce. For this reason a wide array of noninvasive techniques exist whose purpose is to avoid the invasiveness of catheterization. These noninvasive techniques include radionuclide scanning [Gould *et al.*, 1986], magnetocardiography [Geselowitz, 1979], echocardiography [Round, 1982; Wells, 1977], and electrocardiography [Bones, 1981]. Despite continuing advances in these fields, the precise diagnostic tests performed in the catheterization laboratory usually precede surgical treatment. Particularly, angiography is almost the only proven method for examining the abnormalities of coronary arteries [Heintzen, 1986]. This is be-

cause conventional angiography provides information with high spatial, temporal, and density resolution, which on balance is superior to the information from any other modality for imaging the cardiovascular system [Heintzen, 1986].

There are other disadvantages to performing conventional angiography in addition to the invasiveness of catheterization. In order to obtain high quality images of the structure of interest, high dosage X-ray exposure and introduction of large amounts of constrast material are necessary. The use of X-ray exposure and contrast material are known to be potentially hazardous to patients [Maddison, 1977]. Diagnostic decisions from conventional angiographic data are primarily based on qualitative, visual analysis of the passage of a contrast bolus through the circulation. Thus, only information that can be seen instantaneously can be readily incorporated into the decision-making process. Making an accurate diagnosis can be assisted by viewing each angiographic sequence a number of times. However, viewing single frames is less helpful; the cardiac anatomy is harder to perceive when its motion is arrested.

6.1.3 Evolution of Digital Coronary Angiography

With the advent of digital computers, less invasive angiographic techniques for diagnosing circulatory disease have been reported. These techniques, which digitize images obtained from a conventional X-ray imaging system, mainly take advantage of digital image processing techniques for improving the diagnosis in various ways and are commonly referred to as *digital angiography.*

One important stimulus for the development of digital angiography has been the desire to quantify vascular anatomy and function. The quantitative measurement of X-ray densities of contrast-enhanced structures and subsequently the measurement of multiple dimensions of the cardiac chambers or vessels can estimate volume, shape and contraction patterns [Wood *et al.*, 1964; Heintzen, 1971; Heintzen and Bürsch, 1978; Wood *et al.*, 1972; Heintzen *et al.*, 1975] Similarly, there is a need for quantitatively determining the blood flow in vessels by, for example, the indicator dilution method, in which the contrast material is injected into a vessel and its concentration in the flowing blood is measured, thereby providing a measure of the blood flow [Rutishauser *et al.*, 1967; Lantz *et al.*, 1980; Bursch *et al.*, 1981; von Spreckelsen and Wolschendorf, 1985]. For images in video form, the image information is available as an electronic signal, and further analogue or digital electronic processing can be easily performed. Earlier, techniques such as *videodensitometry* or *cinedensitometry* [Wood *et al.*, 1964; Heintzen, 1971; Heintzen and Bürsch, 1978] were originally invoked for accomplishing such tasks, and might be considered the technological predecessors of modern digital angiography.

Another motivation for developing digital imaging methods has been the need to enhance the visualization of specific structures and, in particular, the contrast material used to "indicate" the circulatory structures. Such enhancement seeks to achieve the following goals: imaging with less contrast material (often called dye) and/or radiation with either intravenous or selective intra-arterial injection of contrast bolus [Heintzen, 1986; Harrington *et al.*, 1982; Mistretta *et al.*, 1982; Ratib and Rutishauser, 1985]. It follows that the subtraction technique, originally sug-

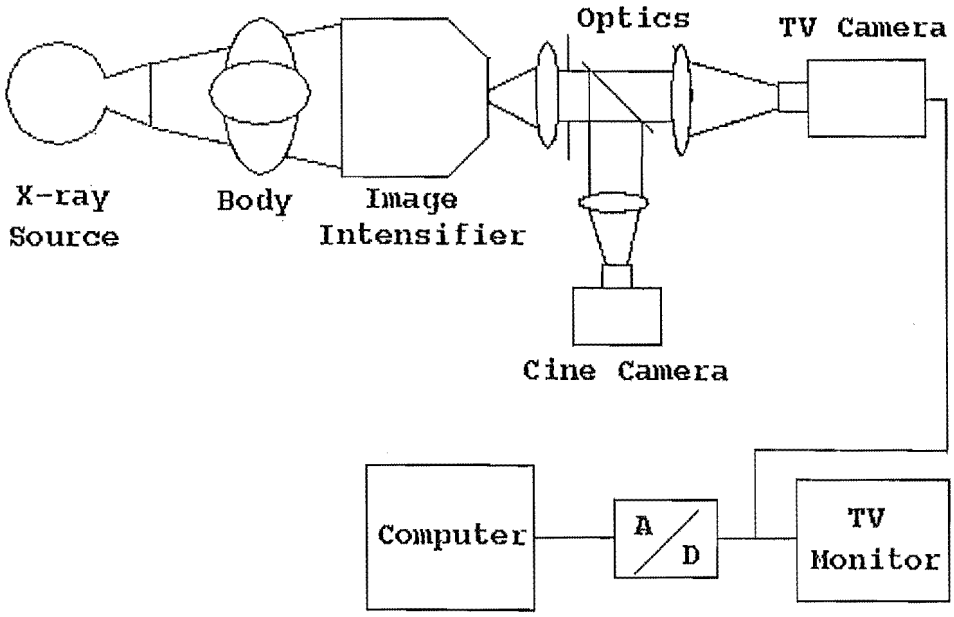


Figure 6.6. Block diagram of a typical digital X-ray angiographic imaging system.

gested by Ziedes des Plantes [1932], is reintroduced into radiographic imaging, but is performed by digital techniques. This is commonly termed digital subtraction angiography (DSA) [Levin *et al.*, 1984; Guthaner *et al.*, 1985; Riederer and Kruger, 1983; Balter *et al.*, 1984], and is widely invoked in clinical practice for imaging the circulatory structure of the body.

A further reason for employing digital technology is the flexibility it can offer for processing, archiving, retrieving, and transferring the massive amounts of clinical data which are routinely recorded [Moores, 1987]. In clinical practice, rapid access to critical data is often a key issue for diagnosis.

6.2 Digital X-ray Imaging System

Digital X-ray angiographic imaging systems (see Fig. 6.6) share the same imaging components that comprise a conventional angiography system, except that images recorded by video cameras are subsequently digitized and processed by digital computers [Kruger, 1986]. An alternative, though clearly less convenient, method of digitizing angiographic images is to record the images on film (of high resolution but limited dynamic range [Collins and Skorton, 1986; Kruger, 1986]) and then use a scanning densitometer to digitize the information recorded on the film.

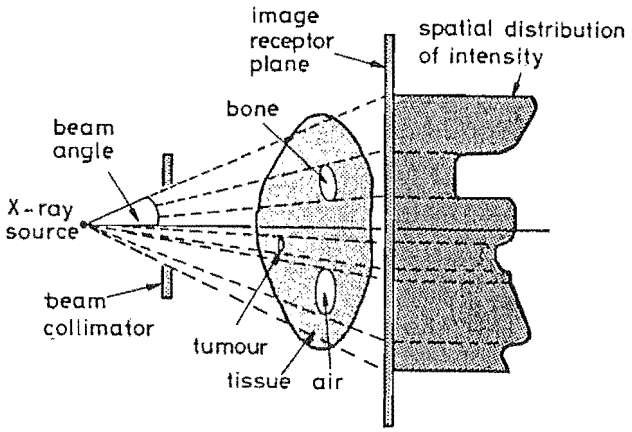


Figure 6.7. Illustration of conventional X-ray attenuation image formation process.

6.2.1 Image Formation

The useful information gathered by almost all X-ray imaging modalities is derived from the attenuation suffered by beams of X-rays passing through a body. The principle of conventional X-ray image formation is illustrated in Fig. 6.7. The mechanism is often analysed by invoking the simplified (or idealised) imaging geometry (known as the *parallel geometry*) which is illustrated in Fig. 6.8 [Macovski, 1983]. Each photon in an X-ray beam either interacts with a particle of matter in the body or passes through unaffected [Johns and Cunningham, 1974]. The interaction results in the removal of the photon from the beam by scattering or absorption. Given that N_0 is the number of photons in an X-ray beam incident upon the body, the following relationship (the Lambert-Beer Law) holds:

$$N = N_0 \exp\left[-\int_0^z \mu(E, Z(x,y,z))\,\rho(x,y,z)\,dz\right] \tag{6.1}$$

where N is the number of photons in the beam remaining after it has passed a distance z through the body, μ is the *mass attenuation coefficient* which depends on the photon energy E and the atomic numbers of, or the charge density $Z(x,y,z)$ of the electrons belonging to, the elements in the material [Johns and Cunningham, 1974], and ρ is the material density.

The number of primary photons, which do not interact with matter and travel out of the body in their original direction, indicate what the attenuation of the body to X-rays actually is because after they have been measured one knows how many photons have not passed through the body. As such, the intensity distribution in the detection plane, i.e. the image $I_d(x,y)$ as indicated in Fig. 6.8, is a 2-D projection of the 3-D body.

In the diagnostic range of X-ray energies, which is below 200 keV [Johns and Cunningham, 1974], there are two main mechanisms governing the attenuation: photoelectric absorption and Compton scattering [Garrett and Smithson, 1987]. Photoelectric absorption occurs when an X-ray photon interacts with and subsequently

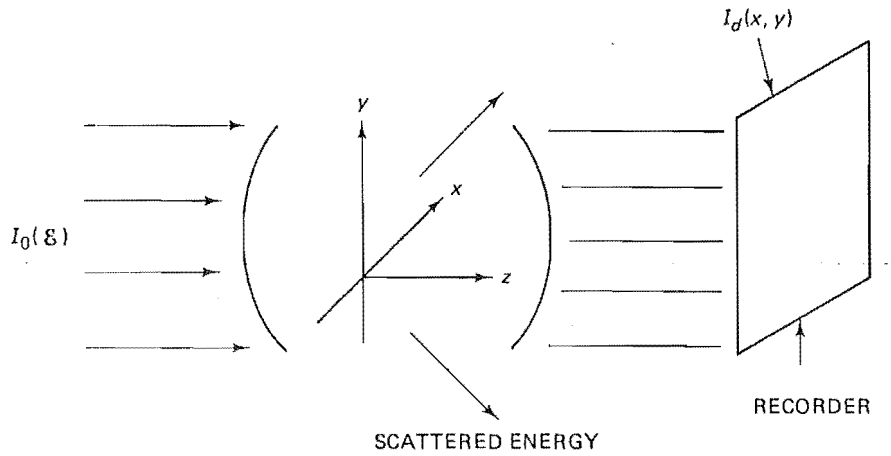


Figure 6.8. Illustration of the idealised conventional X-ray image formation geometry. Note that $I_0(E)$ represents the incident X-ray intensity which is uniform in the x-y plane and $I_d(x, y)$ denotes the intensity of the recorded X-rays, which forms the image.

frees a tightly bound electron in a inner shell. The Compton effect consists of a collision between an X-ray photon and either a free or a loosely bound electron in an outer shell [Macovski, 1983]. The scattered X-ray photon loses a small portion of its energy and travels away from its original direction.

In general, for low atomic number material, such as soft tissue, at low photon energies, attenuation is dominated by photoelectric absorption, while at higher photon energies (above 50 keV) the Compton scattering effect is dominant. For example, the mass attenuation coefficient of water, which is similar to that of soft tissue, is plotted in Fig. 6.9. Low-energy components of the X-ray spectrum (below 50 keV) are largely attenuated inside the human body and play no useful part in image formation [Garrett and Smithson, 1987]. For higher atomic number elements, the photoelectric absorption becomes dominant at all photon energies. This is because the mass attenuation coefficient due to photoelectric absorption varies approximately as the third power of the the atomic number [Macovski, 1983]. Materials of high atomic number elements, such as iodine or lead, highly attenuate X-rays, and therefore can be used as indicators (contrast agent) for imaging vessel structures which are otherwise invisible in an X-ray image. This is because vessels are made of soft tissue, and cannot be discriminated, in an X-ray image, from other parts of the body which are also made of soft tissue. For example, the attenuation coefficient of iodine is shown in Fig. 6.10 as compared to that of muscle and bones. The attenuation coefficient of iodine undergoes a sharp increase in the region of 33 keV, at which energy level many photons interact with the K-shell electrons. This is known as the K absorption edge [Macovski, 1983; Levin, 1984].

Scattered photons (and remember that Compton scattering is the main cause of tissue attenuation in the diagnostic energy region) generally emerge from the the body at a deviated angle (as compared to the incident beam). Unless appropriate precautions are taken, a significant fraction of them are also recorded in the detec-

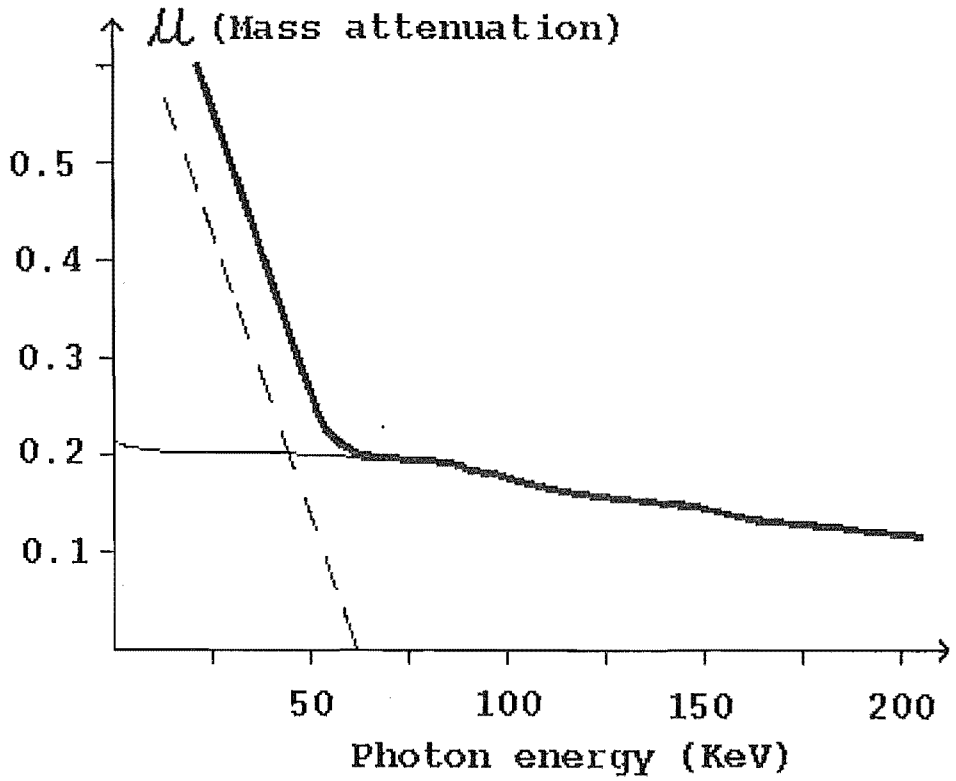


Figure 6.9. Mass attenuation coefficients of water as a function of photon energies: dashed line — photoelectric absorption; solid line — Compton scattering; thick solid line — total mass attenuation coefficient.

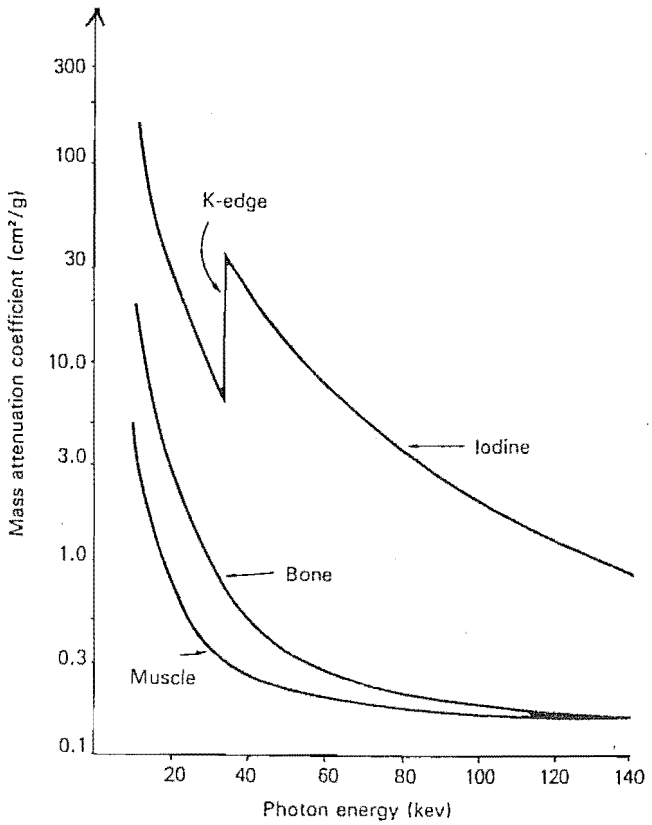


Figure 6.10. Mass attenuation coefficients of iodine, bone and muscle as functions of photon energies. As the photon energy increases, absorption of X-rays by all three substances decreases and does so by different amounts. Note, however, the mass attenuation of the iodine has a sharp increase at 33 keV, which is known as the K absorption edge.

tion plane as primary photons, with consequent degradation of the image. There are two main methods for reducing the number of recorded scattered photons [Macovski, 1983]. One is to interpose a scatter reduction grid (what is often called a collimator in other technical contexts) between the body and the X-ray detectors. The other is to separate the detectors from the body by an appreciable distance so that most scattered photons which travel at deviated angles miss the detector [Macovski, 1983; Garrett and Smithson, 1987]. The latter is seldom desirable since it reduces the signal-to-noise ratio (see §6.2.4) because the average number of photons passing through a unit area per unit time falls off with the square of the distance that the photons travel.

6.2.2 The Imaging Chain

As indicated in Fig. 6.6, X-ray photons emerging from the body are detected by an X-ray image intensifier, which is a fluoroscopic device. With the latest generation tubes, the phosphor scintillator on the input surface converts X-ray quanta into light quanta which, in turn, produce electrons at a photocathode attached to the input phosphor

screen. The emitted electrons from the photocathode are accelerated towards the output phosphor screen while being focussed by the electron optics. The electrons impinging with high kinetic energy on this screen produce light quanta generating a thousand-fold brighter output optical image of reduced size.

The optical image from the image intensifier is coupled to the video (TV) camera using a two-element lens (collimating and objective) optical system. Varying the focal length of the collimating or objective lens affects the actual size of the image formed at the TV camera target. A variable aperture stop is placed between the two lenses to regulate the amount of light that reaches the TV tube, relative to the amount that leaves the output surface of the intensifier. The transmission optics of this system unavoidably generates scattered light that degrades image quality in much the same way as do scattered X-rays. In addition, light scatter and multiple reflections within the transparent window at the output of the image intensifier contribute to light spreading. The combined effect is often called "veiling glare" [Kruger, 1986], which reduces the information that can be abstracted from an image. Veiling glare can be reduced through improved optical design and the use of fibre optic couplings [Kruger, 1986].

All modern TV systems employed in commercially available medical X-ray angiographic equipment employ generic-type vidicon camera tubes [Garrett and Smithson, 1987]. The vidicon may be assumed to consist of three sections [Hill, 1975]: the electron gun, the scanning section, and the target screen made of photoconductive material. The target can be modelled as a very large array of "target elements", each acting as a small capacitor, with one end connected to an electrode (common to all elements) in parallel with a light dependent resistor. In operation, the electron beam from the electron gun is scanned in a raster fashion across the rear surface of the target setting each capacitor to the same potential. When light is focussed onto the target, the conductivities of the elements are determined by the intensities of their respective illuminations. Consequently, each capacitor is partially discharged to an amount proportional to its illumination. Ideally, while scanning the charge pattern, the electron beam re-charges the depleted elements to their original states. In practice, the conductivity of each element decays exponentially once the illumination ceases. This is the origin of the "lag" discussed in Chapter 4. As the raster scan proceeds, current flows through the aforementioned electrode and produces a voltage across a resistor connected to the electrode. This voltage is subsequently amplified, and combined with suitable synchronising pulses, to generate the composite video signal.

Most TV systems (following the European standard) use 625 lines per frame to give good vertical resolution and require 50 images per second to generate effectively flicker-free pictures (this arrangement also accords with the requirement for imaging mobile objects and recording dynamic information). For pictures to have sufficient horizontal resolution a video frequency of about 10 MHz is required [Hill, 1975]. For largely historical reasons, a system known as double interlace is employed to give the same resolution, but using only half the frequency. In this system a video rate of 25 frames per second is employed, with each frame (complete picture scan) comprising two interlaced scan fields. The first vertical scan only covers the "odd" lines, forming

the "odd" field. Next, only the "even" lines are scanned, forming the "even" field. Thus, there are 50 fields per second, each of 312.5 lines.

The video signal must be digitized using an analogue-to-digital (A/D) converter in order to obtain a digital image. At least an 8-bit, or preferably a 10-bit, A/D converter should be employed in a digital angiographic system [Levin *et al.*, 1984] to give sufficient grey level discrimination for the video signals. The maximum allowed quantization (or grey level) spacing should be chosen comparable to the standard deviation of the image noise (see §6.2.4) [Kruger *et al.*, 1981; Bjorkholm *et al.*, 1981]. The digital video samples from the A/D converter are usually dumped to one or more addressable frame buffers (real-time memory) [Bones *et al.*, 1988] hosted in computer memory and subsequently processed by the computer. A minimum resolution of 512 × 512 pixels is required for sampling each video signal frame, for the reasons given in the next subsection.

6.2.3 Resolution Considerations

The resolution of an X-ray imaging system can be no better than that of the component with the poorest resolution. It is important, therefore, to ensure that efforts to improve resolution are undertaken cost-effectively. Improvements in any single component may produce little overall improvement due to limitations elsewhere. In general, major factors which affect the overall system resolution can be grouped into two types: spatial unsharpness and temporal unsharpness [Davison and *et. al*, 1977; Moores, 1984].

Spatial unsharpness is well understood in the context of radiological systems [Macovski, 1983; Moores, 1987; Moores, 1984; Kruger, 1986]. For example, it can be caused by the finite size of the focal spot of the X-ray source [Kruger, 1986; Macovski, 1983] and by object movement during X-ray exposure [Macovski, 1983; Moores, 1984]. Unsharpness also occurs when (photon or electron) quanta interact with a fluorescent screen. At the screen, secondary quanta generated from a single incident quantum spread over a diffuse area. As reading devices, TV cameras have their own spatial resolution limitations. A typical unsharpness of this kind can be described by a point spread function (psf), or by the modulation transfer function (MTF), which is the modulus of the Fourier transform of the psf [Macovski, 1983]. Invoking the concept of spatial frequency, the MTF is a particularly useful measure of unsharpness. For example, the overall MTF of several cascaded (linear) stages is simply the product of the MTFs of the individual stages [Rossmann, 1969].

Temporal unsharpness is caused by the TV camera lag discussed in Chapter 4. Because of the rapid motion of the arteries, the temporal unsharpness represents the single most significant resolution limitation for coronary angiography.

Temporal unsharpness is of course irrelevent when one is imaging static structures of the body. Either the unsharpness due to the focal spot or the resolving power of the TV camera can represent the spatial resolution limit [Kruger, 1986]. It is desirable to use focal spots of very small size for good resolution. However, because one must generate an adequate X-ray intensity and avoid over-heating the target (anode of

X-ray tube), the size of the focal spot can not be decreased indefinitely. The effective focal spot size used in angiography is usually in the range of 1 to $2mm$ [Kruger, 1986]. Under operating conditions usual for cardiac angiography, the distance between the X-ray tube and the detector plane is about $100cm$ and the height of the TV picture corresponds to $150mm$ (the size of heart). Kruger [1986] has shown that the resolving power of a TV camera having 512 vertical lines (of the same order as that of the system described in the last subsection) is similar to that of an X-ray tube with a $1.2mm$ focal spot. If a smaller focal spot is used, or the distance between the X-ray tube and the detector plane is reduced, a higher resolution TV system should be considered. [Kruger, 1986] has also shown on the basis of the sampling theorem that digitization of a video frame into a 512 $\times$ 512 pixel matrix is only marginally sufficient for accommodating the resolving power of the conventional video system described in the last subsection. This matrix size must therefore be considered the minimum that can be accepted.

6.2.4 Noise Considerations

The ability to visualise a structure in a noise-free environment depends on the contrast. Local contrast (within an unit area) can be defined as [Macovski, 1983]

$$C = \frac{\Delta I}{\overline{I}} \tag{6.2}$$

where $\overline{I}$ is the average background intensity and ΔI is the measure of intensity variation within the unit area [Macovski, 1983]. However, contrast is not the fundamental limit on visualization when noise is present. Noise sets the limit on the smallest contrast difference which can be detected. In this case, the signal-to-noise ratio (SNR), a measure of the ability to visualize image detail, is determined by the ratio of the desired intensity variations to the random intensity variation [Macovski, 1983], i.e.

$$\text{SNR} = \frac{\Delta I}{\delta_I} = \frac{C\overline{I}}{\delta_I} \tag{6.3}$$

where δ_I is the rms value of the random fluctuations.

For an X-ray imaging system of the kind described in last section, every stage in the image formation and recording process introduces some noise or statistical uncertainty. The main noise processes are:

1. Quantum uncertainties due to the finite numbers of: (a) photons generated and detected for the first stage of image formation; (b) incident and converted light photons and electrons at each stage of the entire imaging chain [Macovski, 1983].

2. Additive noise due to: (a) Compton scattering and photon scattering during transmission via each optical component of the system [Macovski, 1983]; (b) electronic noise which is added to the video output signal [Moores, 1987; Moores, 1984; Kruger, 1981; Shaw and Bassano, 1982].

The effect of quantum uncertainties on SNR is now considered. It is generally accepted that Poisson statistics [Parzen, 1960] govern both the interaction of X-ray photons with the body and the detection and conversion of quanta at each stage of the entire image chain [Macovski, 1983]. Assume that, on the average, there are N X-ray photons emerging from the body and impinging upon a particular unit area per unit time on the detector surface. Since, in practice, an X-ray photon detector is not ideal, only a fraction of the X-ray photons which impinge on the detector are captured and converted into light photons. The fraction number, denoted by η, which characterizes the X-ray photon capturing ability of the detector is called the *quantum efficiency*. Therefore, the number of X-ray photons that are actually captured is ηN within the unit area mentioned above. This gives an quantum uncertainty (in terms of standard deviation) of $\sqrt{\eta N}$ [Macovski, 1983]. Consequently, the SNR at this stage is, from (6.3),

SEE ERRATA

$$\text{SNR} = \frac{C\bar{I}}{\delta_I} = \frac{C\eta N}{\sqrt{\eta N}} = C\sqrt{\eta N} \tag{6.4}$$

One of the main design criteria for any X-ray imaging system is that image quality must ultimately be limited by the X-ray quantum noise at the first stage of image formation. For the subsequent image formation process (see in §6.2.2) Macovski [1983] has shown that, if the quantum gain (the average number of secondary quanta produced for each input quantum) at each conversion stage is g_i, the overall SNR due to quantum uncertainties is

$$\text{SNR} = \frac{C\sqrt{\eta N}}{\sqrt{1 + 1/g_1 + 1/g_1 g_2 + 1/g_1 g_2 g_3 + \cdots}} \tag{6.5}$$

In practice, g_1, which corresponds to conversion from X-ray to light photons at the fluoroscent screen of the intensifier, is of the order of 10^3, and the gain products ($g_1 g_2, g_1 g_2 g_3$, etc.) range from the order of 10^2 to 10^3 [Macovski, 1983]. It follows that the SNR is dominated by the statistical uncertainty of the captured X-ray photons.

Additive noise due to scattering may be largely reduced by the methods mentioned in §6.2.2. Electronic noise, which is independent of quantum fluctuations, is often termed the system noise. One example is the random noise arising from video signal amplification. Electronic noise is considered to occupy a fraction of the dynamic range of the average value of the video output signal [Macovski, 1983]. It has been shown [Kruger, 1981; Shaw and Bassano, 1982; Moores, 1984] that electronic noise represents the limiting noise component in regions of high photon transmission. Thus, while quantum noise dominates in the dark region of the video image, electronic noise dominates in the white region. An implication is that, as pointed out by Kruger *et al.* [1981], in order to achieve the optimum performance while preventing X-ray exposure from becoming excessive, the maximum X-ray exposure should be chosen to be that for which quantum noise is comparable to system noise.

Electronic noise is bandwidth dependent, varying approximately as $(\Delta f_b)^{3/2}$ (where Δf_b is the bandwidth) [Kruger, 1981]. To satisfy the the sampling theorem, it is necessary that the video bandwidth be the minimum compatible with the video signal digitization rate. For instance, if the digitization rate is 10 MHz (cor-

SEE ERRATA

responding approximately to the order of 500 pixels across the image frame), the bandwidth should be restricted to 5 MHz.

Another noise source is the digitization of the video signal into a finite number of grey levels. This results in an uncertainty in the grey level assignments which is of the order of $\pm 1/2$ of the digital grey level spacing. This spacing must be chosen to match the minimum system noise, and the total number of grey levels must accommodate the dynamic range of the video information [Bjorkholm *et al.*, 1981]. Kruger *et al.* [1981] have shown that 8-bit A/D conversion is adequate for video images formed by X-ray imaging system of the type described previously in this chapter. However, 8-bit A/D conversion must be seen as the minimum requirement for two reasons. Firstly, an X-ray imaging system with particularly low system noise requires a smaller grey level spacing and therefore more quantization levels in total. The second reason concerns the use of averaging. In principle integration of N frames relatively reduces the noise standard deviation by a factor of $\sqrt{N}$. The reduction cannot be achieved in practice, however, if the spacing between two grey levels of the A/D conversion is significantly greater than the noise standard deviation predicted by the averaging process.

In order to study the effect of noise on the phase correlation technique by computer simulation (see Chapter 4), I have simplified the simulation by representing the noise as being both Gaussian and additive. More specifically, computer-generated Gaussian distributed noise was added to the computer-generated arterial images, as described in §4.3.3. The following arguments are offered in justification of this approximation:

1. As argued above, quantum noise can be made comparable in amplitude to the additive system noise;
2. When the average number of detected photons $N >> 1$, the Poisson distribution can be approximated by a Gaussian distribution having an appropriate mean [Feller, 1968];
3. The total noise, as a sum of several noise components, may be loosely described by a Gaussian distribution according to the central limit theorem [Haykin, 1983].

Maher and Thomson [1988] have also demonstrated experimentally that the random noise resulting from subtraction of two images (of same test object) obtained from an digital angiographic system is Gaussian to a useful level of approximation.

SEE ERRATA

SEE ERRATA

6.3 Image Processing Techniques And Motion Problems In Digital Angiography

Over the past 10 years, much research has been conducted into digital processing methods for enhancing visualization of human anatomy and extracting time-related functional information such as blood flow, for both diagnosis and physiological assessment.

Voluntary and involuntary patient motion (for example, cardiac and respiratory motion, swallowing, and any sudden movement [Kruger *et al.*, 1982]) significantly

limits the usefulness of these digital processing methods. This is especially true in coronary angiography [Kruger, 1986; Heintzen, 1986], as is discussed further below.

The principles of the DSA technique are discussed in §6.3.1. After that, in §6.3.2, the basic ideas of blood flow measurement and parametric imaging are introduced. Ways of overcoming the apparent limitations caused by the aforementioned types of motion are also examined.

6.3.1 Contrast Enhancement — DSA techniques

The basic idea underlying the DSA technique is to acquire images before and after a change in the image has occurred due to the passage of contrast material through an organ which is being studied. Subsequently, a difference image is formed in order to enhance the change, and thereby to render the anatomy of the organ more visible. Before going into this any further, a number of terms used in this context are defined:

A *precontrast image* is formed when no contrast medium is present.

A *postcontrast image* is formed at significant (usually maximum) X-ray opacification due to the arrival of contrast material.

Linear subtraction refers to the subtraction of two recorded X-ray images.

Logarithmic subtraction refers to the subtraction of two images which are represented by the logarithms of the recorded X-ray images.

The principle of DSA is illustrated in Fig. 6.11. The top curve in this figure represents an intensity profile taken from a single horizontal line of a precontrast image. The middle curve in Fig. 6.11 shows the same profile from a postcontrast image. The increased X-ray attenuation due to the presence of contrast medium in certain regions of the body results in a decrease of intensity at several positions along the profile. The subtraction of the precontrast profile from the postcontrast profile results, ideally, in only the contrast signal profile being revealed, as shown in the bottom curve of Fig. 6.11. The resulting signal usually has only a small dynamic range. It is usually expanded to match the grey-scale range of an image display device [Collins and Skorton, 1986].

Logarithmic transformation can be performed digitally after the digitization of the detected X-ray image. Alternatively, the transformation can be effected by an analogue logarithmic amplifier before the incoming video signal is digitized. Logarithmic subtraction is often preferred over linear subtraction [Balter *et al.*, 1984] in order that the resulting contrast signal from the subtraction operation depends, roughly, only on the attenuation of the introduced contrast material (see (6.1)).

Major DSA techniques are introduced under each subtitle in what follows:

Temporal Subtraction: In the simplest form of temporal subtraction, the precontrast and postcontrast images discussed above are chosen by visual inspection. This

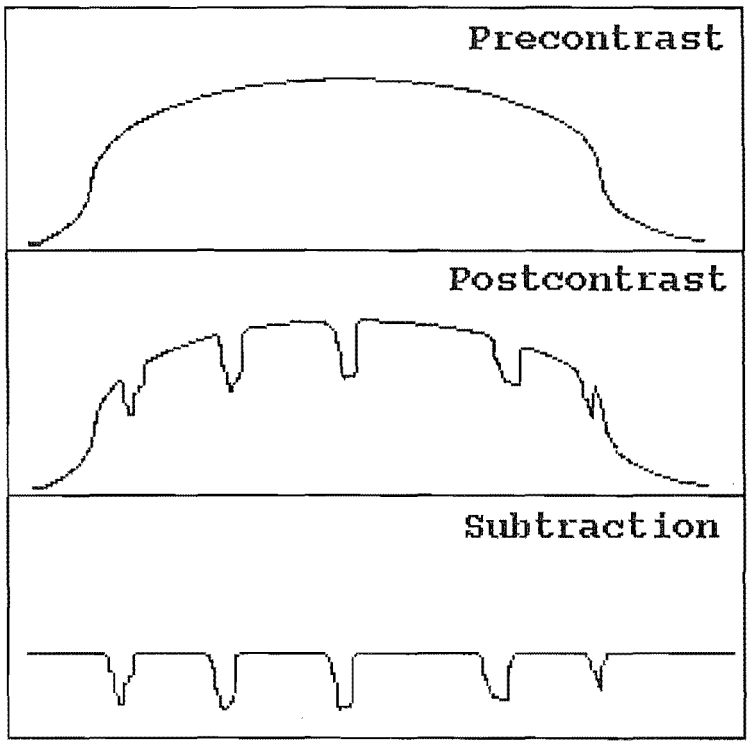


Figure 6.11. Principles of DSA. Top: intensity profile along a single horizontal line from a precontrast image; middle curve: same profile from a postcontrast image; bottom curve: subtraction of the precontrast profile from the postcontrast profile.

is termed mask-mode subtraction where the precontrast image is the "mask" image and the postcontrast image the "live" image [Liu *et al.*, 1985]. The quality of mask-mode subtraction is limited mainly by random noise in situations where motion is insignificant (such as non-cardiac vascular systems) since the noise level in the subtraction image is unavoidably higher than that in either the live or the mask image. Several methods designed mainly for improving the SNR of the subtraction process have been widely studied [Riederer and Kruger, 1983], and are variously called temporal subtraction or temporal filtration. They are characterized in general by a class of temporal linear filters of the form

$$I_s(x,y) = \sum_j k_j I_j(x,y) \tag{6.6}$$

where $I_s(x,y)$ is the resultant image, $I_j(x,y)$ is a frame in a sequence of video images, and j is the instant at which $I_j(x,y)$ is captured. The filter coefficients k_j are constrained by

$$\sum_j k_j = 0 \tag{6.7}$$

to ensure that, ideally, the background is eliminated and only the varying signal due to the introduction of contrast material remains. The determination of these coefficients usually depends largely upon the hemodynamics of contrast medium in blood. More specifically, k_j often depends on the time-varying concentration of the contrast medium in blood as indicated in Fig. 6.12. In temporal filtering, a sequence of images is always acquired to record the complete passage of the contrast medium through the organ being studied (see Fig. 6.12(a) and (b)). Mask-mode subtraction is a special case of temporal filtering as indicated in Fig. 6.12(c). Other forms of temporal subtraction are, referring to Fig. 6.12, integrated mask subtraction (Fig. 6.12(d) [Heintzen, 1986]), matched filtration (Fig. 6.12(e) [Kruger and Liu, 1982]), and recursive filtering (Fig. 6.12(f) [Kruger *et al.*, 1982; Riederer and Kruger, 1983]). Since, in general, temporal filtering involves weighted integration, much higher SNR than that resulting from simple mask-mode subtraction can be achieved [Liu *et al.*, 1985; Riederer *et al.*, 1983].

Temporal filtering techniques have been successful in imaging relatively static vascular structures such as the carotid arteries [Kruger *et al.*, 1981], the aorta and aortic arch [Moodie *et al.*, 1982], and the abdominal aorta and renal arteries [Buonocore *et al.*, 1981], under which circumstances the presence of small motion artefacts does not interfere with obtaining diagnostic information. For imaging highly mobile structures, such as coronary arteries however, temporal subtraction techniques do not usually produce diagnostic quality images. This is particularly true for intravenous injection of contrast medium (see below) since the contrast signal is very weak even with injection of a large amount of contrast material [Heintzen, 1986; Kruger, 1986]. It is worth noting that there is an approach to temporal subtraction DSA, called time-interval-difference (TID) subtraction, which can reduce the influence of respiratory motion [Heintzen, 1986]. In TID each image is subtracted from an image recorded slightly earlier, thus suppressing artefacts due to slow motion [Heintzen, 1986]. TID has been used for extraction of image regions with rapidly changing structures, for instance the left ventricular borders [Kruger *et al.*, 1979].

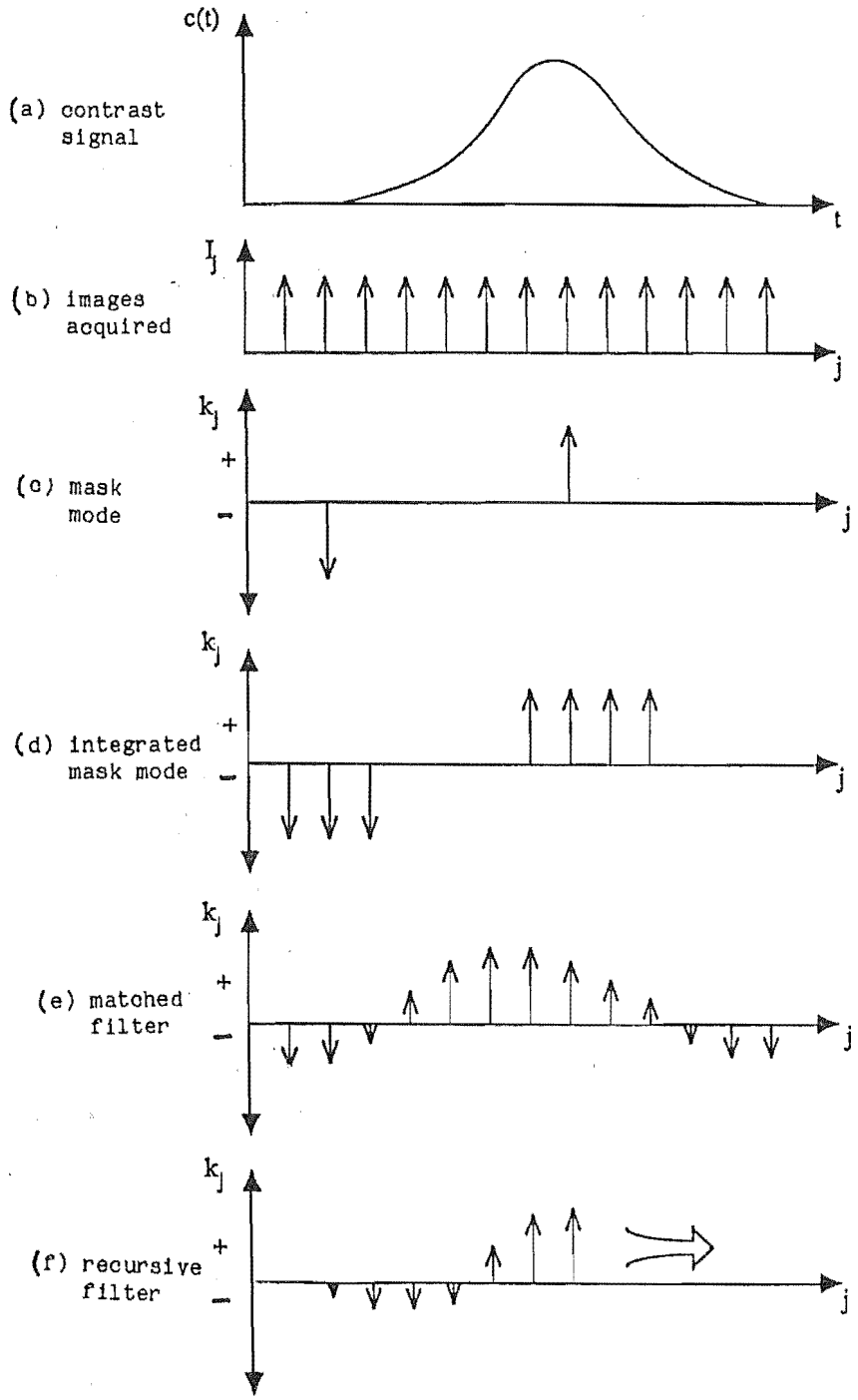


Figure 6.12. Illustration of major types of temporal filtering (or subtraction) technique: (a) *concentration-vs-time* curve (see Fig. 6.13 for illustration of how such curves are formed); (b) acquisition of an image sequence; (c) mask-mode subtraction; (d) integrated mask subtraction; (e) matched filtration; (f) recursive filtering.

Energy Subtraction: Another DSA technique is the so-called dual-energy subtraction [Brody *et al.*, 1981; Kruger *et al.*, 1977; Riederer *et al.*, 1983], in which two images are obtained with different radiation energies and one is subtracted from the other. Referring to Fig. 6.10, consider that one image, denoted as f^{130}, is obtained at 130 keV, while the other, f^{70}, is obtained at 70 keV. In the difference image $(f^{130} - f^{70})$, assuming that f^{130} and f^{70} were acquired at virtually the same instant, the contrast signal due to iodine is relatively enhanced while the soft tissue and bone signals are relatively reduced. In practice the images are acquired as consecutive frames of a video signal and are therefore separated in time by approximately $40msec$. Artefact due to slow motion such as swallowing is largely eliminated [Nelson, 1985]. However, artefact caused by rapidly moving structures, such as the coronary arteries, is not eliminated. Nor are the background structures rendered invisible or the effects of noise avoided [Riederer *et al.*, 1983; Nelson, 1985].

Hybrid Subtraction: Mask-mode and dual-energy subtractions are often used in combination, which is referred to as hybrid subtraction [Brody, 1981; Nelson, 1985]. In this technique, energy subtraction is first performed for a pair of masks and a pair of live images. Next, two of the resultant difference images are temporally subtracted to isolate the contrast signal. Because the energy subtraction step eliminates the soft tissue signal, any tissue motion between the acquisition times of the two pairs hardly affects the final resultant image. Techniques involving energy subtraction involve complicated technical niceties [Nelson, 1985], which is decidedly disadvantageous and restricts the usefulness of such techniques.

ECG and Respiratory Gating: When motion is periodic (or at least repetitive), gating can be employed to reduce motion artefacts. This technique requires the recording and synchronization of image data with appropriate reference signals. For cardiac contraction, images can be obtained at a pre-determined cardiac phase by invoking ECG gating [Heintzen *et al.*, 1978]. The subtraction of two images so obtained reduces motion artefacts due to cardiac motion. Temporal filtering incorporating ECG gating has been successful for imaging heart ventricles [Heintzen, 1986] with intravenous injection of contrast medium. Similarly, respiratory gating [Bogren and Bursch, 1984] can be realised through the use of a respiratory impedance monitoring instrument (e.g. the respiratory inductance plethysmograph [Sinton, 1986]). Subtraction of two images obtained at the same respiratory phase reduces artefacts due to motion of ribs and soft tissues of the chest wall related to respiration [Ro, 1985]. Since cardiac motion and respiratory motion are independent of each other, it is usually necessary for patients to hold their breaths during the whole ECG-gated imaging procedure, which can last for up to 15 seconds, in order to effectively eliminate the influence of respiratory motion [Brennecke *et al.*, 1978; Kruger *et al.*, 1979]. Furthermore, gating techniques generally involve more sophisticated electronic equipment and may therefore not be cost-effective. Ro *et al.* [1987] attempt to reduce both cardiac and respiratory motion artefact by measuring the phases of both cardiac and

respiratory motion for a set of mask images, and then subtracting a synthesised mask image from a live image. However, the results reported for their *computed mask* technique are not very encouraging.

Intravenous DSA and Selective DSA: Intravenously injected contrast material is usually introduced into a peripheral vein of the arm or the leg [Kruger *et al.*, 1979]. Some investigators prefer to introduce a catheter through the antecubital vein into the superior vena cava [Guthaner *et al.*, 1985]. Selective injection, on the other hand, requires that a catheter be introduced into the area being examined through either the venous or the arterial system, depending on the specific clinical situation encountered and on the cardiologist's preference [Guthaner *et al.*, 1985]. The amount of contrast material needed for intravenous injection is much more than that for selective injection in order to achieve sufficient contrast [Heintzen, 1986], but intravenous injection involves less risk and discomfort. Since the introduction of large amounts of contrast material is harmful to the human body, it is an open question whether intravenous DSA is less invasive than selective DSA with reduced amounts of contrast medium. Although intravenous DSA techniques have had great success in imaging, or assessing the state of, the peripheral circulatory and even non-vascular cardiac structures, they have failed in most of their applications to coronary angiography [Ratib and Rutishauser, 1985; Kruger, 1986]. This applies to examining both coronary artery stenoses and the patency of coronary bypass grafts [Guthaner *et al.*, 1985; Heintzen, 1986]. Selective injection techniques and ECG-gated subtraction seem to be the methods of choice for most investigators who currently employ DSA to study coronary circulation.

Further Comments On Difficulties in Coronary Angiography: Researchers (see for example [Heintzen, 1986; Kruger, 1986]) have shown that the digital angiographic techniques discussed above offer many clinical advantages over conventional angiographic techniques. As discussed previously, however, most of those digital techniques are in general sensitive to patient motion, and have been largely restricted to imaging relatively static parts of the body (or those where motion is very small). Gated DSA techniques, especially ECG gating, can partly reduce the influence of motion when imaging the heart. DSA techniques involving frame integration, which is particularly useful for reducing random noise [Huang and Hsu, 1981], suffer strongly from all kinds of motion.

In imaging coronary artery structures, most of those digital techniques fail unless motion is compensated. This is true even with ECG or respiratory-gated imaging. Firstly, coronary arteries of diagnostic interest are usually less then 3 mm in diameter, which makes it difficult to assess degrees of artery narrowing [Kruger, 1986]. Even small motion can cause complete misregistration of the coronary artery in successive images. Secondly, since cardiac motion is not synchronized with respiratory motion, the effects of these two motions cannot be simultaneously reduced unless patients hold their breath. However, even if patients hold their breath perfectly, the ECG gating technique suffers because the heart motion is not precisely periodic. Thirdly,

significant voluntary patient motion is difficult to eliminate, particularly with young patients. It has, therefore, been realised that motion compensation by *pixel shift* is often necessary for applying digital techniques to coronary angiography [Ratib and Rutishauser, 1985]. §7.3 discusses, and presents my recent results on, the application of motion compensation by pixel shift to digital coronary angiography.

To help the reader to appreciate difficulties that can encountered in coronary angiography, I now quote here some comments by Kruger [1986]:

> "... The present and future role of digital angiography in the diagnosis of cardiac disease is not so clear. While non-cardiac digital angiographic equipment proliferates, the introduction of such equipment into the cardiac catheterization laboratory has been slow. The most likely explanation for this difference in level of acceptance between radiologists and cardiologists is the failure of intravenous digital angiography to produce diagnostic quality images of the coronary circulation. While initial attempts at imaging non-cardiac vascular anatomy via an intravenous route of contrast media administration offered great hope, the same imaging approach applied to the coronary arteries has offered only frustration. ..."

6.3.2 Blood Flow and Parametric Imaging — Digital Videodensitometric Techniques

The temporally varying intensity or brightness at a point of interest on a video or film screen can be recorded by a densitometer so as to obtain an *intensity-vs-time curve*. Analysis of such 1-D intensity-vs-time curves is useful for extracting information about the passage of contrast medium through body organs and hence about the function of those organs [Heintzen, 1986; Hohne and Bohm, 1983]. Such techniques are often termed videodensitometry, and they are conveniently implemented by processing digitized video image sequences. A digitized sequence of $N \times N$ pixel images provides N^2 such 1-D intensity-vs-time functions, one per pixel. However, rather than analysing intensity-vs-time curves directly, it is conventional to convert such curves into the so-called *pixel-densograms*, which are in fact the 1-D concentration-vs-time curves of contrast medium mixed in blood. The conversion from each intensity-vs-time curve to each pixel-densogram is illustrated in Fig. 6.13. The pixel-densogram at pixel spatial position (i, j) can be expressed as $C(i, j, t)$, where t is time. Various time-related parameters can be extracted from the analysis of densograms [Kruger *et al.*, 1983; Heintzen, 1986; Hohne and Bohm, 1983]. Some examples are indicated in Fig. 6.13(c). The application of these extracted parameters to angiographic examinations is twofold: blood flow measurement and parametric or functional imaging, both of which are briefly discussed in what follows.

Blood flow Measurement: Blood flow measurement is important for the study of human circulation. It is often performed during and following cardiac surgical

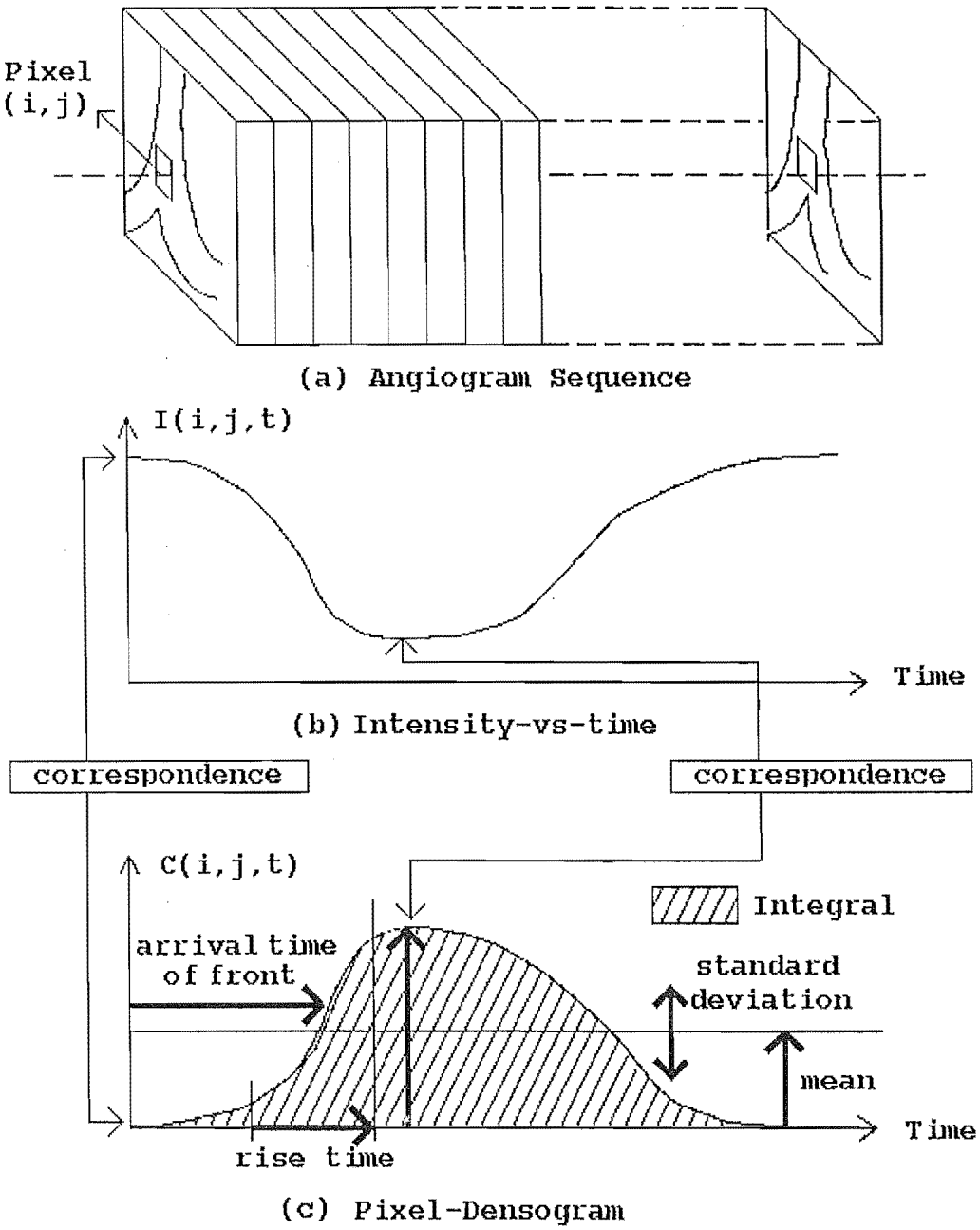


Figure 6.13. Illustration of the formation of pixel-densograms: (a) a digital angiogram sequence; (b) intensity-vs-time curve; (c) pixel-densogram (i.e. concentration-vs-time curve). Several useful parameters for angiographic studies can be extracted from the pixel-densogram. They are are indicated by black arrows and a hatched area. Each arrow represents a particular parameter, and the length of the arrow indicates the magnitude of the parameter.

operations [Smith *et al.*, 1973; Ratib and Rutishauser, 1985; Vogel *et al.*, 1984; Heintzen, 1986]. Comparison of blood flow between normal and stenosed vessels permits assessment of the severity of the stenosis [Ratib and Rutishauser, 1985]. The coronary flow reserve, defined as the ratio of hyperaemic to resting blood flow, is an excellent indicator of how significantly a coronary lesion affects the perfusion of non-infarcted myocardium [Vogel *et al.*, 1984]. There are two main methods of measuring blood flow in practice [von Spreckelsen and Wolschendorf, 1985]. The first, often termed the indicator-dilution technique, was suggested by Lantz *et al.* [1980] (see also Bursch etal 1981, Vogel etal 1984). It allows the determination of relative blood flow at a particular location in a vessel with respect to a reference location upstream in the circulation.

A technique for estimating absolute blood flow, often termed the *transit-time* method, was first suggested by Rutishauser *et al.* [1967] (see also [Kruger *et al.*, 1983; Smith *et al.*, 1973]). The principle of this method is now described in some detail to indicate how parameters characterizing the transition of the contrast medium in a vessel can be extracted from pixel-densograms. With reference to Fig. 6.14, densograms A and B (shown in Figs. 6.14(b) and 6.14(c)) correspond to points A and B located within a segment of artery in the angiogram (Fig. 6.14(a)). The transit time Δt is the difference between arrival times of maximum opacification by contrast material at point A and point B, i.e. $\Delta t = t_B - t_A$. When the average diameter $\overline{d}$ and the central line length Δs of the segment between A and B are calculated, the average blood flow of this segment is [Rutishauser *et al.*, 1967]

$$Q = \pi \frac{\Delta s}{\Delta t} (\frac{\overline{d}}{2})^2 \tag{6.8}$$

Edge detection may be invoked to derive $\overline{d}$ and Δs [Suntheralingam, 1987]. Relative flow can also be obtained by comparing the calculated flow with that at a reference point.

As for DSA techniques, the success of densitometric techniques is limited by the effects of motion. If the vessel is in motion during the recording of the image sequence, the shape of densograms obtained will deviate from their true forms. A coronary artery vessel is thin and may move a distance of from 1 to 5 times its diameter [Smith *et al.*, 1973]. Therefore, instead of the densogram at only one pixel point, the average densogram for a window centered at the point is usually invoked (see Fig. 6.14(a)). The window may be chosen to be sufficiently large so that the artery segment remains within its border throughout the sequence. However, the densogram for a large window is likely to reflect not only the concentration within the target artery but also that within neighbouring structures. Motion compensation of an artery segment by pixel shift, as described in Chapter 4, allows an optimal window size to be used so as to obtain densograms in which the above-mentioned deviation is effectively minimal.

Parametric or Functional Imaging: Parametric imaging [Hohne and Bohm, 1983; Hunter *et al.*, 1986; Heintzen, 1986] enables visualization of dynamic information

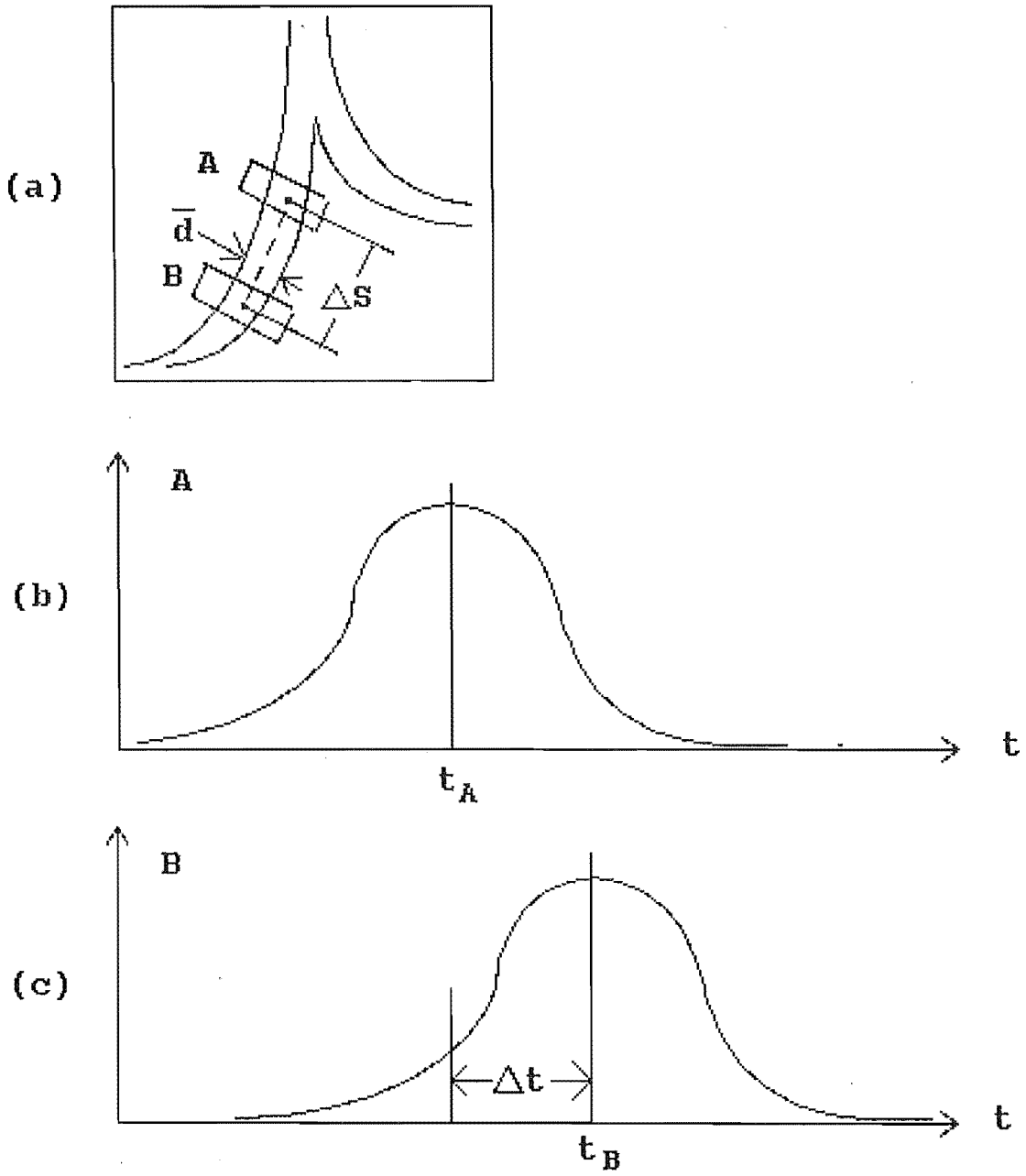


Figure 6.14. Blood flow measurement by transit-time method: (a) an "angiogram"; (b) densogram A corresponds to point A in the angiogram shown in (a); (c) densogram B corresponds to point B in the angiogram shown in (a).

(which is originally contained in a sequence of images) in a single image. A parametric image is of the same size as original images in the sequence, and displays the spatial variation of one parameter (for example, one of the parameters illustrated in Fig. 6.13(c)) in either grey scale or a colour code. A parametric image is obtained by assigning each of its pixels the prechosen parameter value which characterizes a corresponding pixel-densogram of the image sequence.

Time-related parameters, for example the arrival time of maximum or 50% contrast opacification [Heintzen, 1986] (see also Fig. 6.13(c)), are employed for showing regional delay of blood flow which may be caused by certain disease, for example arterial stenoses [Hohne and Bohm, 1983]. They can also be invoked for depicting the blood flow pattern by grouping parameter values into so-called *time segments* [Heintzen, 1986].

Relative blood flow distribution can also be visualized in a parametric image [Bursch *et al.*, 1981]. This is often done by colour or (grey scale) coding each segmental flow (the flow of a very short vessel segment) and labelling it with a percentage number with respect to a reference vessel segment where maximal flow is measured.

Parametric imaging can also be invoked for enhancement of the contrast signal [Heintzen, 1986]. In DSA techniques, for example mask-mode subtraction, contrast signal for all regions can not be equally enhanced since maximum opacification will not be reached simultaneously for all regions in the live image. If the value of maximum opacification for each densogram is displayed in a single image, irrespective of the time when it occurs, contrast signal is equally enhanced for all regions. The parameters, such as the standard deviation, mean value, integral (area under the densogram), etc. (see Fig. 6.13(c)) can also be invoked for this purpose [Hohne and Bohm, 1983; Hunter *et al.*, 1986].

Motion affects parametric imaging in the same way as it affects blood flow measurement, because the first step of both procedures is to form densograms. Some parametric imaging may involve studying a relatively large area of circulation, for example an image of the whole brain or a complete coronary arterial tree. Consider the situation of imaging a complete coronary arterial tree. Since the arterial tree is not a rigid object, the motion of the heart causes different parts of the artery move differently. This makes it necessary to compensate for the motion of each part of the tree independently. While many parts effectively translate, others rotate or even change their shapes. It is not yet clear to what degree local motion compensation by pixel shifting can improve the usefulness of parametric imaging for a complete coronary arterial tree.

Chapter 7

The Cardiology Angiographic Processing System (CAPS) at PMH — Results of Processing Coronary Angiograms

In this chapter, the work performed in conjunction with the Cardiology Department, Princess Margaret Hospital (PMH), Christchurch, New Zealand, is described. First, in §7.1, a specialist hardware and software package, known by the acronym CAPS, is introduced. Next, the method used to remove a particular form of interference associated with angiographic image acquisition, "spike" noise, is discussed in §7.2. Finally, in §7.3, results of motion estimation and compensation for actual CA image sequences are presented.

7.1 The Development of CAPS

Several commercial DSA systems with the ability to perform cardiac imaging are available, but their high capital cost [Evens, 1983] makes them difficult to justify for a facility performing only cardiac investigations. Furthermore, for coronary studies, these systems need to have additional facilities (software) to perform motion analysis. However, conventional cineangiographic systems which, in addition to performing film angiography, record video onto tape, have been widely implemented in major hospitals. Also, currently, a large number of plug-in image acquisition/display boards are available for the major bus families with the least expensive board (for the IBM PC family) costing in the vicinity of NZ$3,000. By replaying the video tape, a recorded angiographic sequence can be input to such an image board to be subsequently digitized and processed by the host computer. This latter approach is taken for studies carried out at the Cardiology Department, PMH, and is considered appropriate at a research level.

The Cardiology Angiographic Processing System (CAPS) has been established at PMH [Bones *et al.*, 1988]. The principal author of this system is Dr. P. J. Bones (now

a Senior Lecturer in Electrical & Electronic Engineering at the University of Canterbury, formerly a Senior Fellow of the New Zealand Heart Foundation). I, together with a former M.E. student [Suntheralingam, 1987], have contributed to CAPS on aspects of digital coronary angiography. CAPS, as a software package, is hosted in an Intel 310 microcomputer which is sited next to the catheterization laboratory where a CGR cineangiographic system (made by THOMSON-CGR Medical Corporation), is situated. The CGR system employs a Thompson CSF vidicon camera, and analogue images are recorded on 3/4" U-matic tape for subsequent analysis. The images are simultaneously recorded on 35 mm cine-film. From tape the video signal is input to a Matrox MIP-512 video processing board which is capable of digitising video frames at a resolution of 512 $\times$ 512 $\times$ 8-bits and at a sampling frequency of 10 MHz. As part of the CAPS software, low-level procedures written in $PL/M-86$ are used to control the operation of the board and manage the transfer of data via a MULTIBUS I backplane, a 16-bit bi-directional microcomputer bus, on and off Winchester disk. CAPS has the ability to "grab" images from a standard ($50Hz$, 625-line interlaced) video signal in real-time into one or more 512 $\times$ 512 $\times$ 8-bits memory-mapped frame buffers, and look-up tables to transform incoming and outgoing data. In practice, five such frame buffers are used in CAPS.

One major feature of CAPS is the video-rate processing of CA image sequences [Bones *et al.*, 1988]. Image sequence acquisition is prerequisite for motion estimation. The facilities for image sequence acquisition and display allow sequences of images to be acquired. In general, the smaller the image size, the more frames can be obtained in the image sequence. For most studies, image sequences of up to either 20 plus frames of 100 $\times$ 100 pixels images, or 12 frames of 200 $\times$ 200 pixel images are captured. More details concerning the components of CAPS and their operation are seen in Appendices B and C which are two articles [Wu and Bones, 1987; Bones *et al.*, 1988] that I have co-authored [see also: Suntheralingam 1987]. Appendix C mainly addresses how conventional image acquisition hardware can be invoked for video-rate image processing.

I emphasise that the majority of my research work during my Ph.D study was not conducted with CAPS but in the VAX/VMS system of the Department of Electrical and Electronic Engineering, University of Canterbury. Clinical images were obtained using CAPS and transferred to the VMS system for subsequent studies. Some software developed under VMS during these studies was then implemented on CAPS.

7.2 Spike Interference and Its Removal

During routine cineangiographic procedures, the circuitry of the CGR cine motor control unfortunately generates significant interference to the video signal. The interference is not serious enough to impede the routine procedures, but it degrades the quality of individual video frames digitized by CAPS. The main component of the interference is distinctive short-duration spikes. Each spike is similar and distinct from the underlying video image intensity variation.

At first glance, it may seem appropriate to eliminate such interference at the source. However, this would have made it necessary to modify the cineangiographic system and interfere with the routine protocol of the investigation. So software was developed to remove the interference by image processing. Studies showed that spikes were triangular and appeared in bursts with a repetition rate of approximately 2 KHz (corresponding on the average to one spike every 14 image rows, i.e. lines of digitized video signal). While the spikes exhibited some variation in shape and height they were reasonably uniform [Wu and Bones, 1987]. Based on the nature of the spikes, a spike detection and removal algorithm was developed. The detection part of this algorithm is mainly based on the principle of the "matched filter", and, after detection, the spike is removed by interpolation from neighbouring rows. This algorithm has been found to remove the spikes successfully without introducing any noticable artefacts, thereby permitting subsequent digital angiographic studies to be carried out essentially free of the interference. The reader is referred to Appendix B for a more comprehensive description of this algorithm.

7.3 Motion Estimation and Compensation of Coronary Arteries

One of the major concerns in digital coronary angiography is the assessment of narrowing of the coronary artery. There are two objectives: 1) to confirm the existence of narrowing; 2) to assess the degree of narrowing of a coronary artery segment by, for example, calculating the diameter of the narrowed artery lumen. The second objective is crucial when deciding whether or not an operation is needed to bypass the narrowed segment with a coronary graft. The qualities of the original X-ray images of the coronary artery are usually poor, and in particular, the blur caused by the lag and the granulous nature of the photon noise (see all figures in this section) prevents the edges of a narrow artery lumen being clearly seen. One can expect appreciable errors if such images are employed directly for computing the diameter of an artery lumen, especially of a narrowed part (current clinical practice still relies on human operators to track the edges of the artery lumen subjectively). Therefore, image enhancement naturally becomes a highly desirable preprocessing step. The algorithm for achieving such enhancement is discussed in §4.4.1, and is illustrated in this section with their applications to four sequences of CA images. Each sequence relates to a practical situation illustrating either the quality of the captured images or the complexity of motion that the coronary arteries undergo.

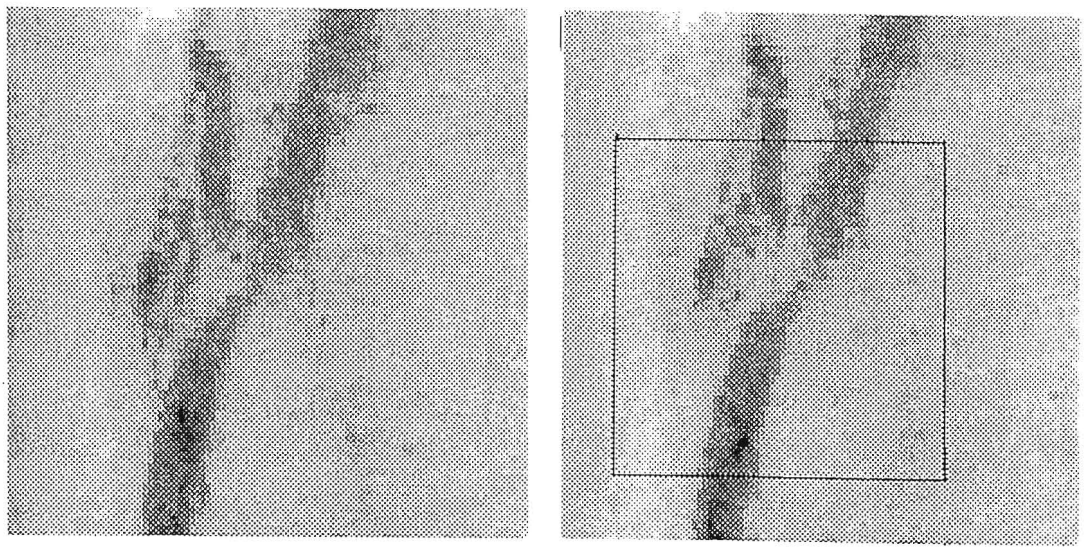
The first sequence which is shown in Fig. 7.1 consists of ten 100 $\times$ 100 pixel CA images. In order to display the FOI more clearly, the linear background is removed in this sequence (see §4.3.1). Refer back to Fig. 4.4(a) for a reminder of what one original frame looks like without the background removed. Note that Fig. 7.1, which highlights the area containing the AOI, only spans about 70% of the area of Fig. 4.4(a). These angiograms show that a narrowing is likely to have occurred near the bifurcation of the artery apparent in the images. Because of the blur in the vertical direction caused by the lag effect, and also because of the image

noise, the narrowing is somewhat obscurred in some frames, and therefore needs to be further confirmed by appropriate averaging procedure which compensates for the motion. Such can be successfully effected by the procedures introduced here.

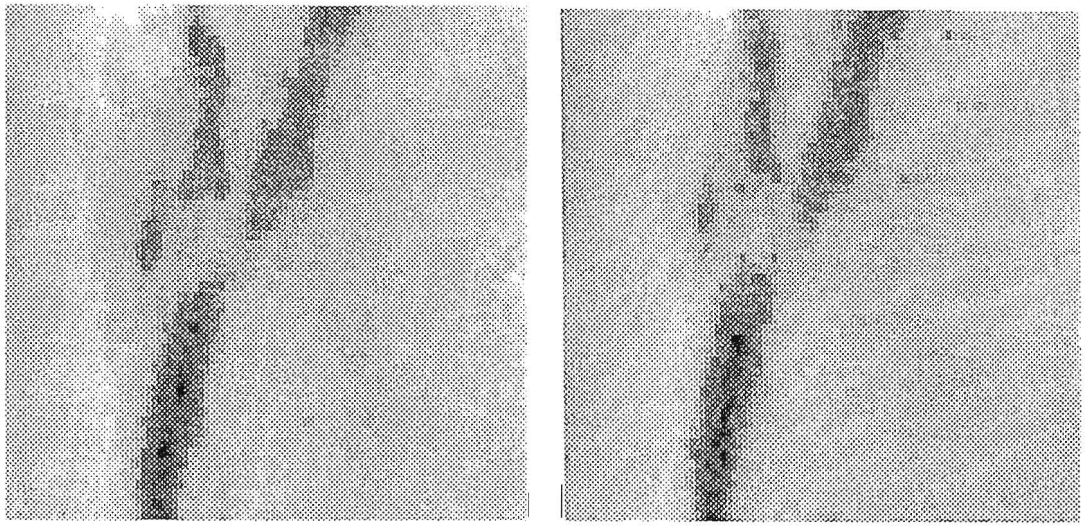
Simply by visual inspection, one sees that the quality of the images is reasonably high, in terms of how clearly the contrast medium is revealed over the background. One also sees that the artery moves by roughly 20 pixels vertically in the sequence (causing blur in the vertical direction) and that the motion is almost purely translational. In this case, the algorithm presented in §4.4.1 can be invoked straightforwardly. The AOI (64 $\times$ 64 pixels) chosen for the second frame is highlighted in Fig. 7.1(b). Remember that the lag correcting step in the algorithm results in only 9 frames, because the first frame is discarded. Taking the bottom left corner of Fig. 7.1(b) to be the image space origin, the coordinates of the bottom left corner of the AOI is $\mathbf{P}_1 = (10, 12)$ (in pixels). This is chosen to be the initial AOI position for the algorithm (see the algorithm in §4.4.1). The success of the algorithm applied to this sequence is illustrated in Fig. 7.2.

Fig. 7.2(a) shows the AOI, which is highlighted in Fig. 7.1(b). Fig. 7.2(b) shows the result of correcting the camera lag in the AOI, i.e. the result of subtracting Fig. 7.1(a) from Fig. 7.1(b), for $k = 0.6$. Fig. 7.2(c) and (d) show the phase difference and phase correlation between Fig. 7.2(b) and the AOI of the third frame in the sequence shown in Fig. 7.1, after the later has been corrected for lag. Note that the appearance of a sawtooth phase (introduced in §4.3.2) in Fig. 7.2(c), although largely corrupted, can still be noticed, indicating that the useful information about the translation of the FOI remains in the phase difference. The position of the peak identified by an arrow in Fig. 7.2(d) indicates the estimated shift $\hat{\mathbf{s}}_{3,2}$, as defined in (4.13), of the FOI between the second frame and the third. After performing the algorithm presented in §4.4.1, the average of the AOI with shift correction, which is shown in Fig. 7.2(e), is generated. Note that Fig. 7.2(e) is the average of the images of the AOI from frames 2 to 10, because the first frame is discarded in the lag correcting step of the algorithm. For comparison, the average without shift correction is shown in Fig. 7.2(f). While both averaged images show reduced noise levels, the motion has clearly blurred the features of the artery in Fig. 7.2(f). Fig. 7.2(e) clearly reveals the narrowed part of the FOI, and confirms that narrowing has occurred. Note that almost the same result is obtained when a variant of the algorithm, which computes for the absolute shift (AS) (see §4.2.1 and §4.4.1) instead of the RS, is applied to this sequence. The reason for this is that the motion of the FOI is almost purely translational.

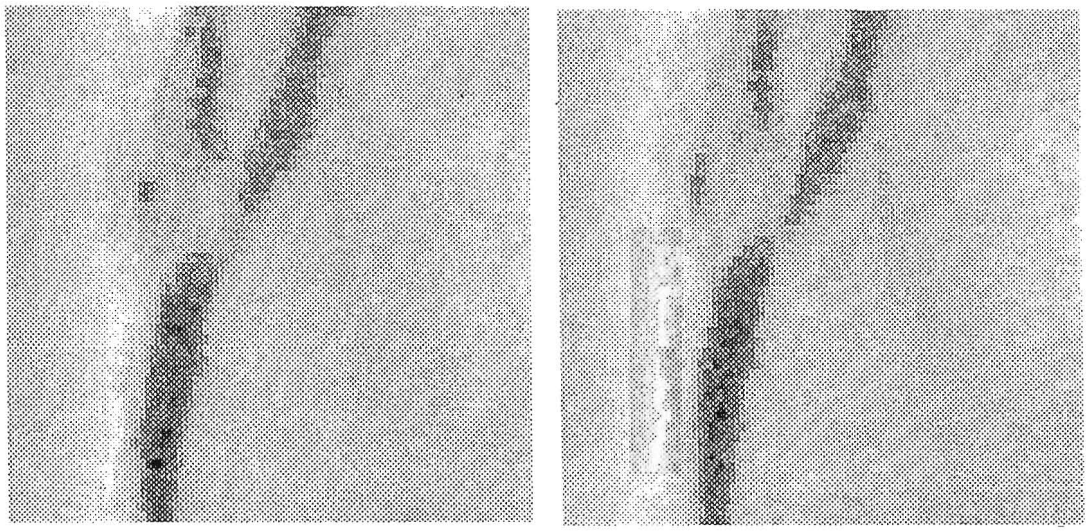
The second sequence consists of twelve 60 $\times$ 60 pixel images of a segment of a right coronary graft. The first and the last frame are shown in, respectively, Fig. 7.3(a) and (b). Some electrical interference has obviously occurred while acquiring this sequence, causing apparent "band" like noise in the images. It can be seen, however, that because the segment was not moving quickly when the sequence was captured, the blur caused by the camera lag is not very significant, as is confirmed from inspection of Fig. 7.3(a) and (b). In fact, it is pointed out in §4.2.2 and §4.4.1 that the lag correction step in the algorithm may be safely ignored (another variant of the algorithm). The RS computation scheme is again invoked for processing this sequence. An AOI of



(a) (b)



(c) (d)



(e) (f)

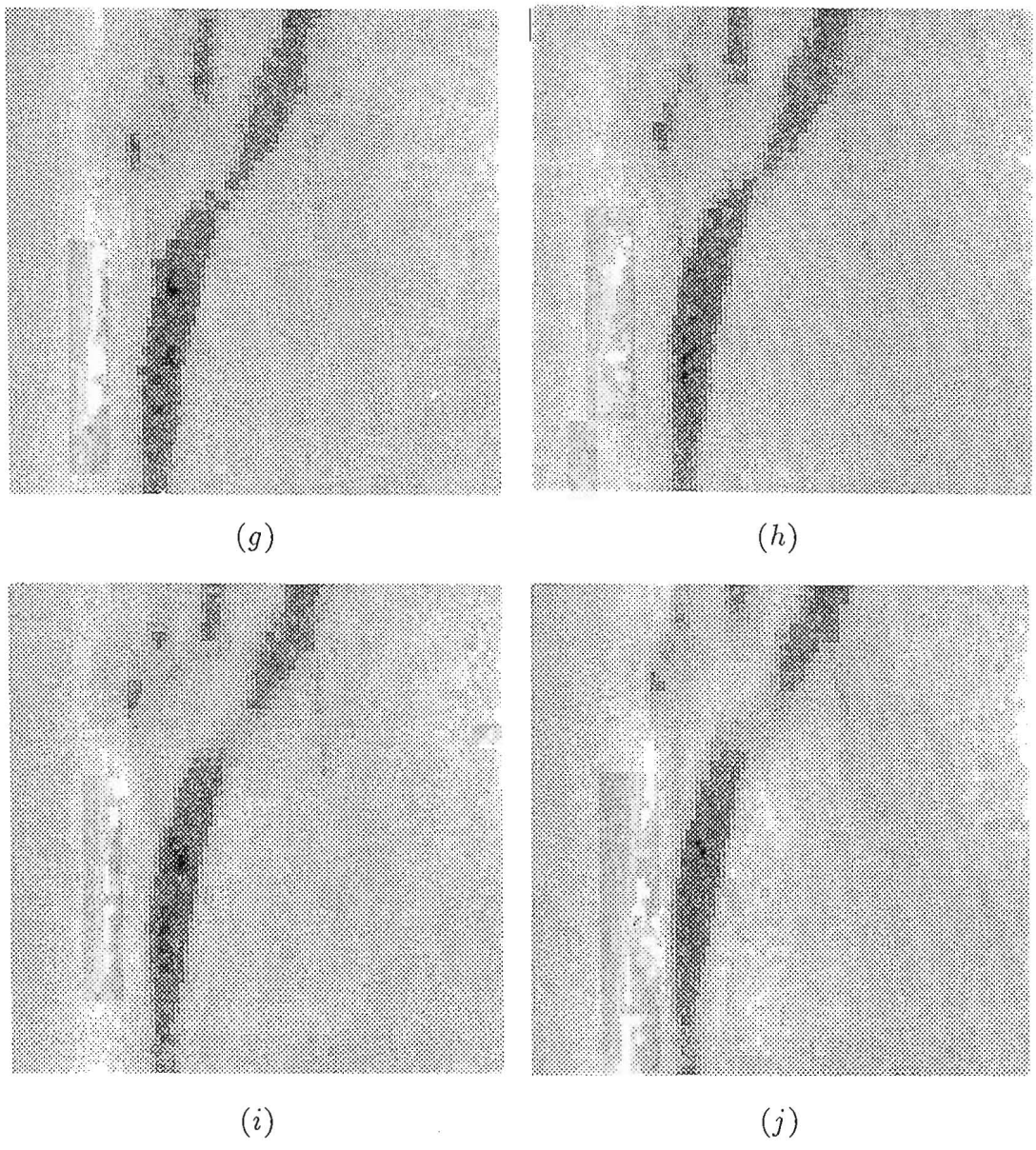
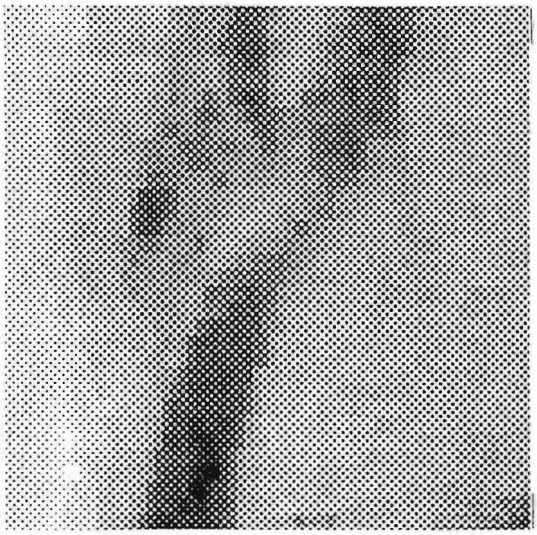
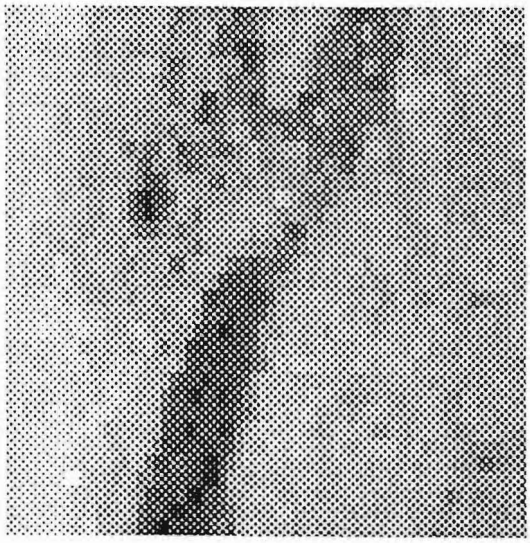


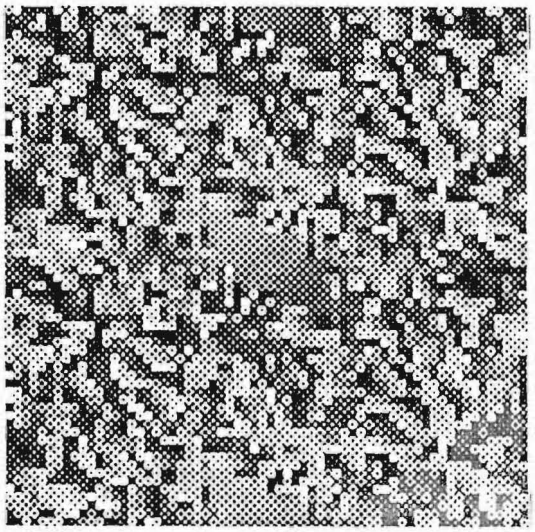
Figure 7.1. A sequence ((a), (b), (c), $\cdots$, and (j)) of 10 consecutive frames (note that these 100 $\times$ 100 pixel images are reduced as compared with the original image Fig. 4.4(a), for convenience of presentation) of the right coronary artery. The coordinates of the bottom left corner of the AOI (64 $\times$ 64 pixels), i.e. the initial AOI position $\mathbf{P}_1$, in the second frame are chosen to be $(10, 12)$ from the bottom left corner of (b), as shown highlighted in (b). The maximum vertical shift is found by SPPC to be 26 pixels from (b) to (j). Because the artery moves significantly in the vertical direction, the blur in the vertical direction is greater than that in the horizontal direction.



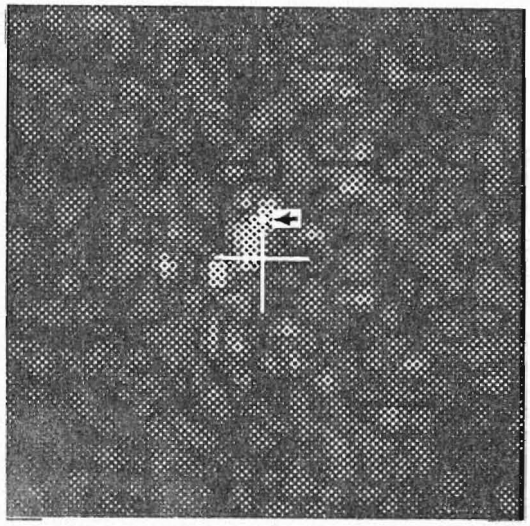
(a)



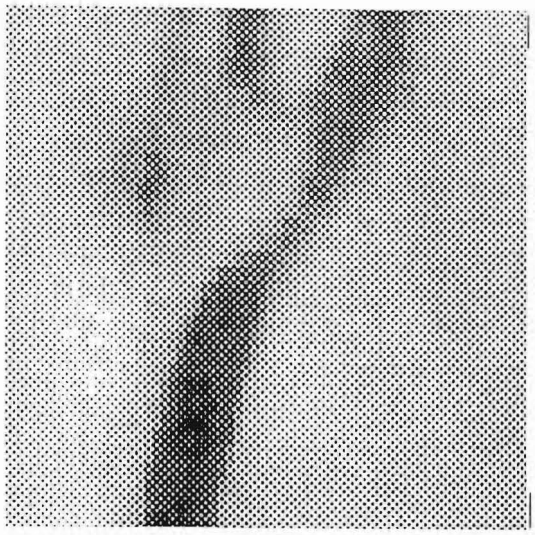
(b)



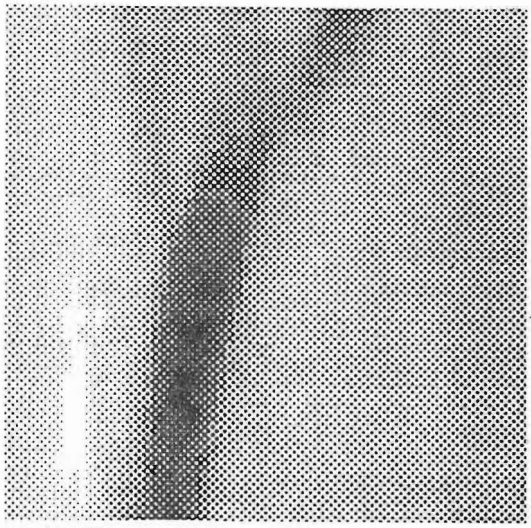
(c)



(d)



(e)



(f)

Figure 7.2. Motion estimation and compensation for the artery segment shown in Fig. 7.1, with averaging of the AOI in all images of the sequence with shift correction to generate an enhanced FOI: (a) AOI extracted from Fig. 7.1(b); (b) the lag corrected image of the AOI, i.e. result of subtracting Fig. 7.1(a) from Fig. 7.1(b) for $k = 0.6$; (c) and (d) phase difference and phase correlation corresponding, respectively, to (a) and to the AOI of the third frame shown in Fig. 7.1(c) after correction for the camera lag; (e) average of 9 AOIs with correction for shifts as described in the algorithm presented in §4.4.1; (f) average of 9 frames without correction for the significant shifts of the artery segment.

only 32 × 32 pixels is used and found adequate for this sequence. The AOI for the first frame is shown outlined in Fig. 7.3(a). To illustrate the result of the motion estimation, Fig. 7.3(c) and (d) show, respectively, the averages of the 12 frames without and with shift correction. It is also concluded, by analogy (and in accord) with the example illustrated in Fig. 7.2, that while both Fig. 7.3(c) and (d) show an improved SNR, Fig. 7.3(c) is noticeably blurred because of misregistration.

The third sequence consists of ten 100 × 100 pixel images of a segment of a left coronary artery. It is seen from Fig. 7.4(a) and (b), which are the second and the tenth frames, that the noise level in this image sequence is high, and the contrast between the contrast medium signal indicating the arterial segment and the background is weak. Two short and separate arterial lumen segments are seen to lie along a common axis in each frame of the sequence. It may be deduced that these two segments belong to the same artery, and that a narrowing has possibly occurred between them. Fig. 7.4(c) shows the second frame after applying the lag correction with $k = 0.55$. It is clear that the contrast medium signal is almost completely overwhelmed by the noise. In this case, the algorithm presented in §4.4.1 is invoked with an additional step of median filtering immediately after the lag correction. An AOI of 64 × 64 pixels is used, as highlighted in Fig. 7.4(a). Corresponding to this AOI, Fig. 7.4(d) illustrates the result of averaging, after lag removal, median filtering and shift correction, 9 frames (frames 2 to 10), which markedly improves the SNR. Furthermore, the apparent narrowing of the artery is clearly revealed in Fig. 7.4(d). One can even discern details of narrowed segment, such as the gradual reduction in the diameter of the artery. This contrasts with what seems to be abrupt blockage in the original images.

The final sequence (ten 100 × 100 pixels images) discussed here is illustrated in Fig. 7.5. Only the result of averaging is discussed here. Some frames of the original CA images (the third, fourth, ninth, and a lag corrected version of the fourth frame) of the sequence are displayed in Fig. 4.3. This sequence displays the most significant blur (see Fig. 4.3) which is caused by the camera lag. This is clearly revealed because of the rapid movement of the artery (inspect frames 3, 4 and 9 of this sequence shown in Fig. 4.3). By visually examining the sequence, it is apparent that the motion involves a slight rotation and, possibly, some change in arterial shape. Fig. 4.3(c) indicates that a narrowing may also have occurred at the position of the bifurcation. It is appropriate to invoke the algorithm presented in §4.4.1 in this situation. The final average with shift correction is shown in Fig. 7.5. While the narrowing is more clearly revealed, the detail of the arterial bifurcation on the right of the narrowed

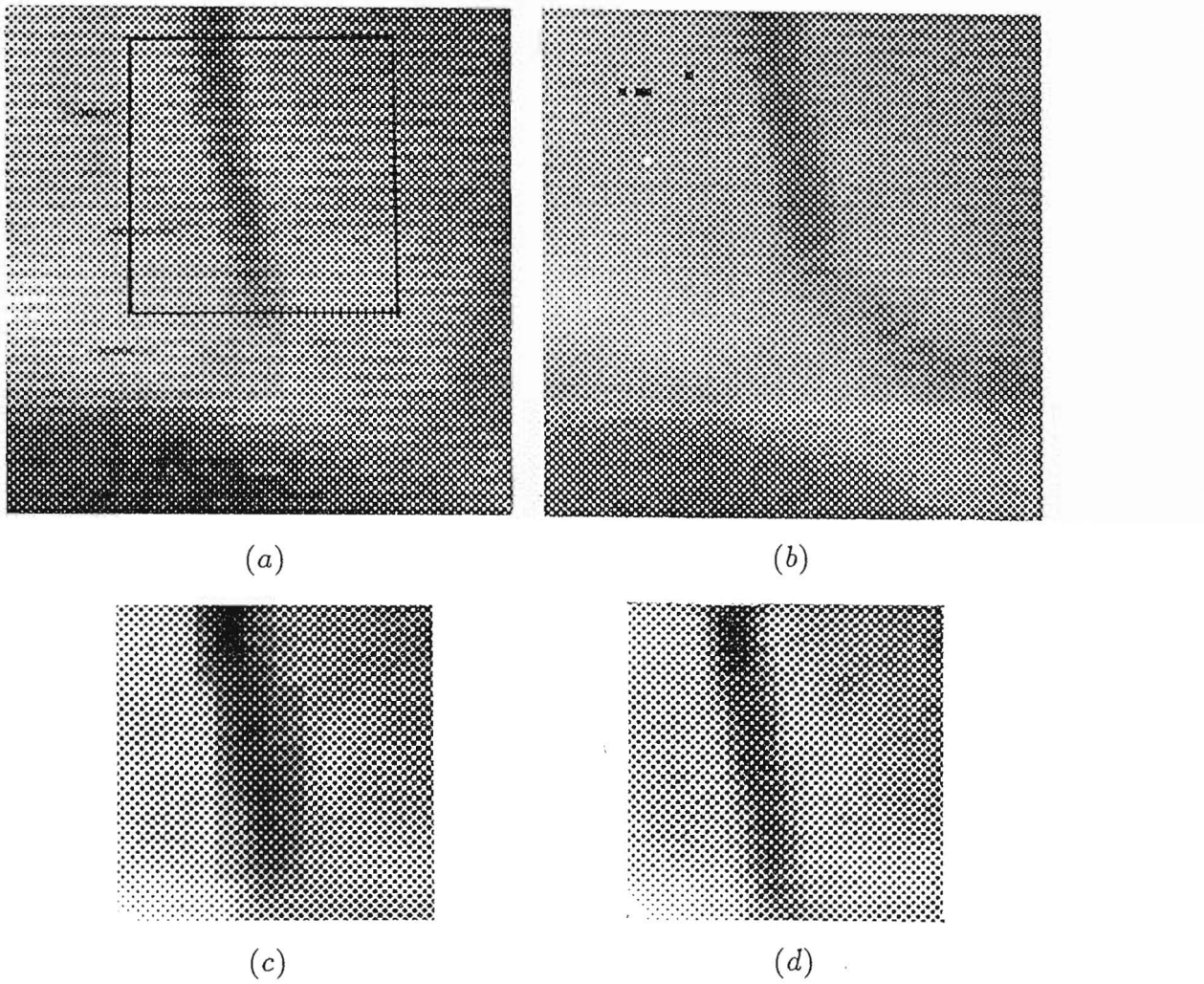


Figure 7.3. Motion estimation and compensation for a segment of a right coronary graft. The sequence consists of twelve 60 × 60 pixel images. (a) 1st frame, wherein the AOI is highlighted; (b) final frame; (c) average without shift correction; (d) average with shift correction. Note that the blur due to the lag effect is not significant so that the original images can be used for motion estimation and frame integration.

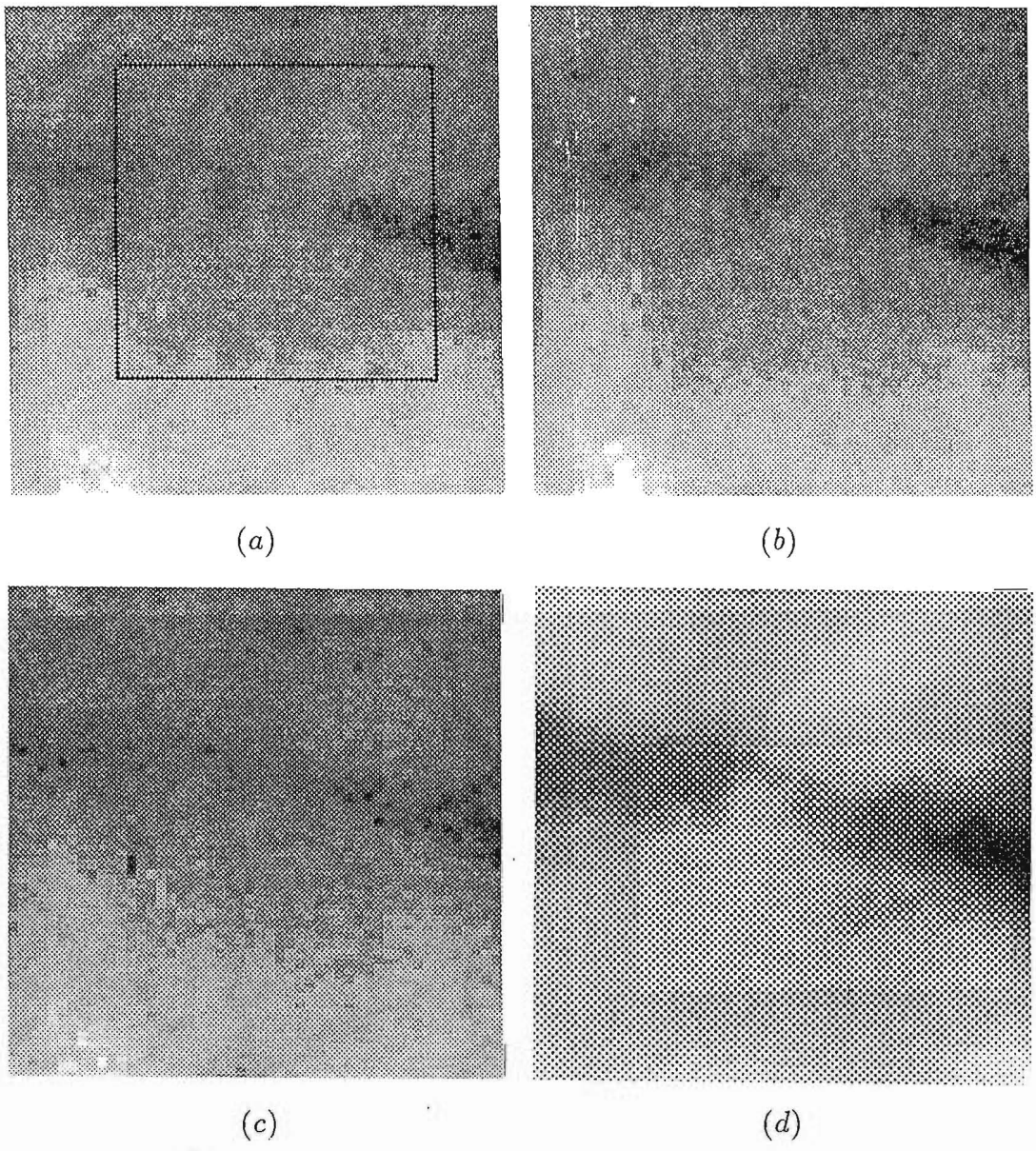


Figure 7.4. Motion estimation and compensation for a segment of a left coronary artery. The sequence consists of of ten 100 × 100 pixel images. (a) second frame, with the AOI (64 × 64 pixels) shown highlighted; (b) final frame; (c) second frame after lag correction with $k = 0.55$; (d) averaged AOI with shift correction, showing detail of the narrowed part which is invisible in the original CA images. Note that after lag correction, the images are median filtered before motion estimation and frame integration are effected.

part is still as obscure as that discernable in Fig. 4.3(c), and less apparent than that in Fig. 4.3(d), which is how the fourth frame appears after the camera lag is removed. The deformation, i.e. the rotation of the artery and its change of shape, may have affected the averaging step of the algorithm. This is because that the final averaging step in the algorithm discussed in §4.4.1 is based on the assumption that the FOI is a rigid object and translates only. Although phase correlation is insensitive to slight rotations when it estimates translations, as argued in §4.2.2, all slight rotations between successive frames estimated by SPPC can add up so as to blur the FOI in the averaging step.

SEE ERRATA

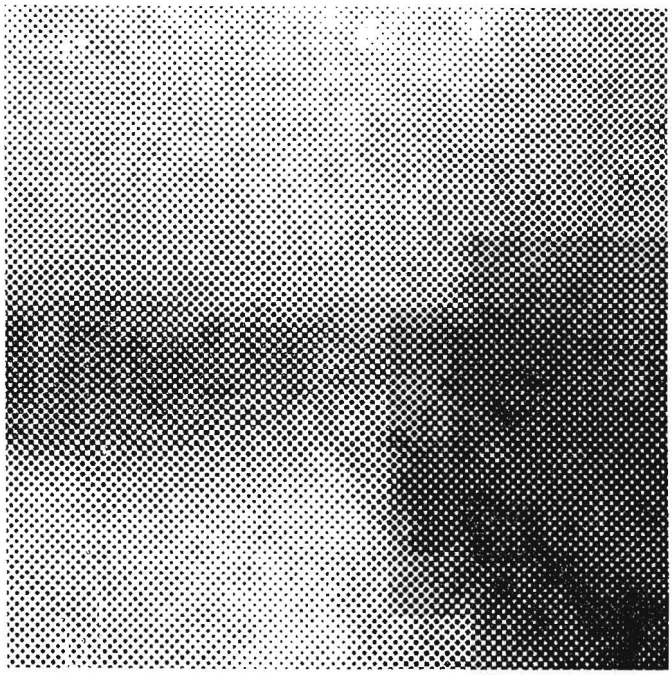


Figure 7.5. Motion estimation and compensation of a sequence of right coronary angiograms (ten 100×100 pixels images depicted in Fig. 4.3): Average of the AOI from frames 2 to 10.

Chapter 8

Conclusions and Suggestions For Future Research

Original (computational and theoretical) results on motion estimation and its application to coronary angiography based on image (sequence) processing techniques are reported in this thesis. This chapter begins by summarising (in §8.1.1) previous work, then discusses and assesses (in §8.1.2) the significance of the major topics examined in this thesis with emphasis on my original work, and finally concludes by suggesting (in §8.2) future research which might usefully be undertaken to extend the results presented in this thesis.

8.1 Discussions and Conclusions

8.1.1 Summary of Previous Work

Image sequence analysis and processing is a recently established research area evolving from many separate research areas such as medical imaging, machine vision, geoscientific image analysis, etc., that is, virtually any area where dynamic scene analysis is involved. Motion estimation, a term broadly referring to estimating any scene changes, is the key issue [Huang, 1981; Nagel, 1983; Aggarwal and Martin, 1983] for dynamic scene analysis (see §1.1). Motion problems are considered to belong to three groups: single object motion, multiple object motion, and 3-D motion (§1.2.1), in order of increasing difficulty. 2-D motion (translation and rotation) is the main concern of this thesis. Since many techniques exist for straightforward estimation of translational motion, the important considerations are the relative accuracies and computational efficiencies of the techniques. Rotation estimation has been much less studied because of, primarily, computational complexity and expense. Methods for motion estimation can be primarily classified into spatial domain and spectral domain approaches. Spatial domain techniques can be further subdivided, e.g. intensity based matching, feature based matching, and the method of differentials (§1.2). As far as 2-D motion is concerned, matching techniques are preferred to the differentials method because the former are less restrictive than the latter. Spatial domain

techniques have found widespread applications in various areas (§1.2).

Spectral domain techniques (Chapter 2) have been relatively less explored than spatial domain techniques. The major advantage of spectral techniques is that translation and rotation estimation can be performed separately, so that motion problems can be simplified, with consequent significant reduction in computational cost. Previous work has mainly been based on conventional second-order spectral analysis. Methods for estimating object translation between two of its images, relying on spectral analysis, can be divided into three approaches (cf. §2.3), based on, respectively, phase differences, the Fourier transform magnitude of the difference image, and the cepstrum. The latter two have only been proposed recently, and have not been discussed by many researchers. Phase correlation (§2.3.4), which is classified as a phase difference method, has been considered to be an alternative to the most popular method, namely matching based on conventional cross correlation. The main advantage of phase correlation over conventional cross correlation is that it can estimate shifts more accurately, because the phase correlation peak tends to be narrower than that of the cross correlation function (§2.3.4). Comparison with the cross correlation technique also shows that phase correlation is less demanding computationally (§2.3.4).

Digital coronary angiography is discussed in detail in Chapter 6. The heart's anatomy and the function of the coronary arteries are described. Coronary disease
 and and its treatment are reviewed. Angiography has been the dominant approach
 for diagnosing coronary abnormalies. Digital technology has been introduced into angiography in recent years due to the advent of digital computers and image processing. Digital technology has brought fundamental changes to conventional angiography. Not only has the quality and quantity of diagnosis been improved, but also many new studies, especially of physiological or functional aspects, have been made possible. Most modern digital angiography studies rely on conventional X-ray and video systems to acquire image sequences. Physical aspects of such imaging systems are reviewed, providing a background to facilitate the understanding of image noise and of resolution limitations. It is claimed in this thesis that most existing digital angiographic imaging techniques can be classified under two categories: image enhancement and functional imaging. These techniques have mostly been predicated on the assumption that images contain static objects, or objects which are in regular motion (e.g. periodic motion). Digital angiography has not been as successful when applied to the coronary system as when it has been applied to other parts of the human circulation. The primary reason is that the motion of the heart causes the coronary arteries to move in a very complicated fashion, thereby contradicting the above assumptions. Motion estimation has to be performed before applying enhancement or functional imaging techniques to images of coronary arteries.

8.1.2 Assessment of Original Work

One significant aspect of the original work reported in this thesis is my approach to phase correlation (PC). The characteristics of PC are related to the concept of phase dominance in §2.3.3. It is shown that PC is closely related to matching of

object features since the feature information is contained in the phase of the Fourier transform of the object (§2.3.4). Phase correlation applied to highly textured images is shown to be more advantageous than it is to less textured images (§4.3.2). Furthermore, windowing in image space causes insignificant degradation to the phase information (see §4.3.2) because object features persist after operating on the image with a smooth window. Three modifications (background removal, image windowing and noise filtering) to the original PC technique are shown to be effective and necessary for applying PC to real world images, which of course exhibit various imperfections (see §4.3). Background removal reduces the intensity variation within an area of interest (AOI) and relatively enhances the feature of interest (FOI), permitting phase correlation to estimate the motion of the FOI without significant bias (because of the existence of significant background intensity variation in most images of the AOI). Windowing the image reduces the corruption of phase information caused by truncation. Phase correlation performed on noisy images usually produces a number of adjacent sharp peaks of similar height in correlation space. If noise filtering is carried out before performing phase correlation, ambiguity in determining the peak corresponding to the object translation tends to be significantly reduced.

What is also significant is the establishment of the sequential predictive phase correlation technique (SPPC). SPPC allows one to process, automatically, image sequences for motion estimation. The original phase correlation can only estimate shift between two images when the shift is small in comparison with the size of the images. Within an image sequence, an object (the FOI) can be enclosed within an area of interest (AOI). When the FOI moves, SPPC predicts the position of the AOI in successive frames of the sequence. SPPC can thus estimate object translation so long as the edges of the images do not truncate the FOI significantly. Only after applying SPPC to coronary angiographic sequences was I able to achieve consistent results.

Like cross correlation, both the PC and SPPC methods could be readily and inexpensively implemented in hardware provided the image size (number of pixels) is not very large, the most demanding operation being the 2-dimensional FFT. The multiplication of the input images by the Hanning window and the filtering of noise are operations which could be performed efficiently by array-processing circuitry. Implementing SPPC on images with a 64×64 pixel AOI by invoking state-of-the-art integrated circuitry could, one has reason to believe, process sequences of images at a rate of several per second. At present, the implementation of SPPC on a VAX 3500 (programmed in Pascal and Fortran) takes approximately 6 seconds to process a frame of a sequence of 100×100 pixel images each containing a 64×64 pixel AOI.

The second aspect of my original work which is commented on here is my method of compensating for limitations in the X-ray/video imaging system (see §4.1 and §6.2.2). The brightness persistence of the video camera screen, the lag, caused by the imperfect photoconductivity of the camera, is known to decay exponentially. The effect of the lag on an X-ray/video system is that each image obtained from the system is effectively a multiple exposure, implying that the image of a moving object is thereby significantly blurred. A weighted subtraction technique is developed in §4.1.2 for compensating for this lag. The method can achieve a resolution comparable

to that of the images of a static object. However, the SNR of the lag corrected image is much lower than that of each original image. It is shown that, when the object translates uniformly, the lag has no effect on phase correlation (§4.2.2). It is also shown that, if the lag factor is less than 0.6, phase correlation can still be performed usefully for arbitrary translation motion (§4.4.2). It is further shown (in §7.3) that the SNR can be significantly improved by appropriate averaging after motion compensation.

I have also designed an image sequence processing algorithm based on the SPPC and the lag correction techniques. I discuss both it and a few of its variants in §4.4.1. Significant new results are obtained from applying the algorithm and its variants to coronary angiogram sequences (§7.3). These results show that these algorithms can estimate coronary artery motion to a useful accuracy. There are two immediate clinical applications of these results: to confirm the existence, and to assess the degree, of coronary narrowing resulting from coronary disease (cf. §6.1.1).

Insight is provided into the degree to which the noise in a complete digital coronary angiographic imaging system can be usually assumed to be Gaussian. I discuss the clinical significance of arterial motion estimation, especially with regard to the techniques for enhancing coronary arterial segments developed in this thesis. It is also suggested that motion estimation can be expected to improve functional (or parametric) imaging, including blood flow measurement.

Another significant contribution reported in this thesis is the application of the third order spectrum, the bispectrum, to motion estimation (Chapters 3 and 5). The theoretical background for motion analysis, in particular the shift and rotation invariant properties (§3.2.2) of the bispectrum, are presented in Chapter 3. It is shown that the complex bispectrum is shift invariant, in contrast to the conventional spectrum, of which, only the magnitude is shift invariant. Based on the concept of the congruent triangle described in §3.2.2, the bispectral angular correlation technique is established in §3.3. The major advantage of bispectral analysis over analysis based on the spectral magnitude is that phase information can be incorporated into an algorithm for estimating rotational motion. This significantly sharpens the peak of the angular correlation, which then becomes a sensitive rotation estimator (§5.2). The computational results presented in Chapter 5 confirm that, because of this advantage, rotation estimation by angular correlation employing the bispectrum is much superior to that employing the conventional spectral magnitude in situations where the level of noise and other degradations is appreciable.

8.2 Suggestions For Future Research

8.2.1 High Resolution Estimation

It would be worthwhile to make an extensive study of possible improvements to the SPPC motion estimation method, which could be realised by incorporating into it the subpixel registration technique described in §2.3.2 and §2.3.4. I found that an estimation resolution of a single pixel seemed to be sufficient for my digital coro-

nary angiographic studies since the image quality of the coronary angiographic (CA) images available to me was inherently poor. It may, however, be worth attempting greater sophistication for coronary angiographic studies based on the most advanced of currently available angiographic equipment. Subpixel estimation will require a significant increase in computational effort because of the need for resampling images to implement both the motion estimation procedures and the subsequent enhancement processing.

8.2.2 Possible Applications of Bispectral Angular Correlation To Real-world Images

One possible future use for the the bispectral approach described in Chapters 3 and 5 is its application to digital coronary angiography. It is shown in §7.3 that, although the motion of a selected coronary arterial segment is usually predominantly translational, small rotational motions cannot be ignored if one is to enhance the segment by averaging a sequence of images. Therefore, a possible approach would be to incorporate the bispectral angular correlation (into the algorithm presented in §4.4.1 for estimating the rotational motion of an arterial segment) in such a way that it operates before the translational motion is estimated by phase correlation. This may very well significantly reduce the blurring of the averaged image caused by the rotation. This may not be an easy task because one has to find out whether or not the bispectral angular correlation is capable of dealing with images which, as remarked in Chapter 4, exhibit few strong features but are nevertheless replete with real world imperfections. One key issue will be how difficult it is to find an optimal congruent triangle, of the kind described in §5.2, for the simple structure of an arterial segment embedded in a complicated background. It is likely to be worth applying background removal, windowing in the image domain and noise filtering (see §4.3) before attempting to estimate rotation. However, it is as yet an open question how effective these preprocessing steps will be in practice.

Another possible application of the bispectral approach is to pattern recognition in industrial automation contexts. Suppose that machine parts (for example, tools, nuts and bolts, etc.) travel on a conveyor belt through an assembly department of a factory. Suppose further that a certain part is to be picked up by a robot hand. In this case, a video camera (the vision system of the robot) can be positioned above the belt in such a way that each frame it captures is dominated by the image of a single object, surrounded by other objects which are truncated by the frame. The object can be arbitrarily oriented, but it must lie in a stable mechanical position on the conveyor belt, implying that very few (perhaps only one in most cases) views of each object need to be pre-stored as templates. The image captured by the vision system could then be processed by the computer incorporated into the vision system. After the object travels further down the belt it may be picked up by the robot hand, which will then know exactly where the object is and how it is oriented.

In order to recognise a particular object, a template must be generated for each object that could be seen on the belt. A descriptor for any particular object can be generated from an image of the object recorded previously. To be as effective

as possible, a bispectral descriptor for an object must correspond to an optimal congruent triangle in the object's spectrum.

The following matching procedure might be implemented. First, the Fourier transform of a captured image is computed. Then, for each descriptor and its corresponding congruent triangle, the bispectrum of the captured image corresponding to that triangle is computed, and is angularly correlated with the descriptor. Subsequently, the position of the major peak in each angular correlation is taken to indicate a possible orientation (relative rotation with respect to the descriptor) of a particular object for which its descriptor is invoked. The determination of what the object is and its orientation in the captured (or the input) image might be as follows: for each possible relative rotation resulting from correlating the bispectrum of the image with a particular descriptor, a rotation compensated image of the input is generated; then each object for which a descriptor is generated is compared to a corresponding rotation compensated image. Various criteria must be devised for determining the best match. For example, the SAVD criterion is very simple to invoke (see §1.2.2). More conveniently, the rotation compensation and subsequently the comparison could be performed for each bispectral descriptor.

As an example, consider determining the orientation of a nut surrounded by other objects. Fig. 8.1(a) shows the nut, captured by the CCD (charge-coupled device) TV camera system set up in the Electrical & Electronic Engineering Department of the University of Canterbury, and subsequently hard-limited in software, such that the intensity of the recorded image, denoted for convenience by $f(\mathbf{x})$, is 1 or 0 depending upon whether the captured intensity does or does not exceed a specified threshold. The hard-limited image, denoted (again for convenience) by $g(\mathbf{x})$, shown in Fig. 8.1(b) reveals the nut shifted (by approximately 3 pixels both horizontally and vertically) and rotated (by 45°) and surrounded by other objects (as they might appear on a conveyor belt). Note that the other objects are truncated at the edge of the finite frame which is a portion of the field of view of the TV camera system. One would expect to be able to estimate by how much the nut has rotated by performing a bispectral angular correlation. An appropriate bispectral descriptor of the nut can be constructed from the triangle shown highlighted in Fig. 8.1(c), which displays $|F(\mathbf{u})|$.

The bispectral descriptor $F_{\rho,\tau,R}(\phi)$ is shown in the top curve of Fig. 8.2, corresponding to the triangle shown highlighted in Fig. 8.1(c). The part of the bispectrum $G_{\rho,\tau,R}(\phi)$ of $g(\mathbf{x})$ (i.e. the image shown in Fig. 8.1(b)), corresponding to the same triangle, is displayed in the middle curve of Fig. 8.2. Because of the presence of other objects in $g(\mathbf{x})$, $G_{\rho,\tau,R}(\phi)$ is considerably different from $F_{\rho,\tau,R}(\phi)$ in addition to being rotated, as confirmed by comparing them with each other. The angular correlation $\Gamma(F_{\rho,\tau,R}, G_{\rho,\tau,R}, \phi)$ is shown in the bottom curve of Fig. 8.2, the major peak of which occurs for $\phi = 44°$, which represents a negligible error for the majority of practical robotic applications. On the other hand, as shown in the bottom curve of Fig. 8.3, the rotation estimate obtained by performing spectral magnitude angular correlation is $\phi = 57°$, which represents an unacceptable error for most practical applications. It is evident that the bispectral descriptor is markedly superior to the spectral magnitude descriptor for this example. The primary reason for the marked contrast between the

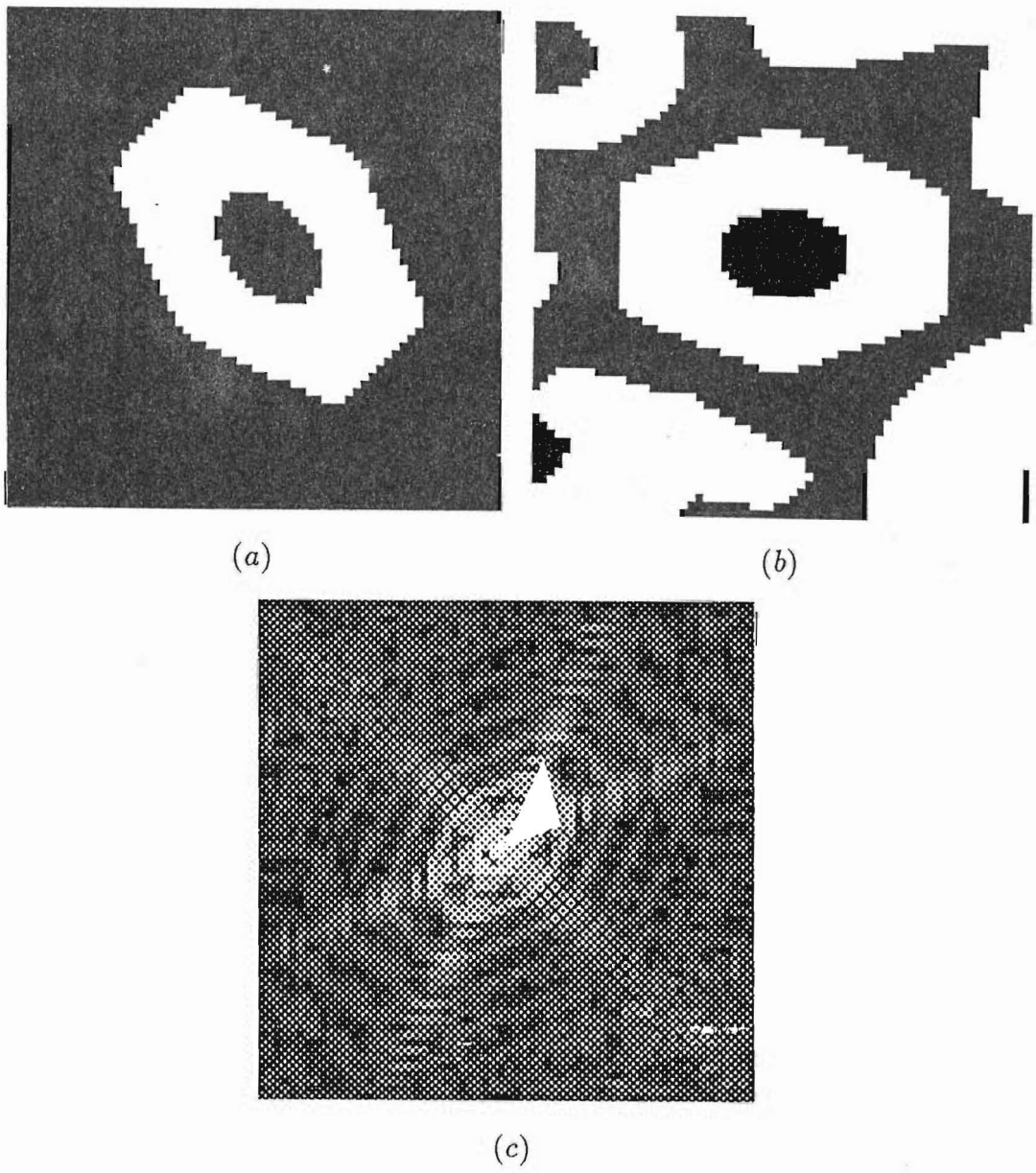


Figure 8.1. Images of a nut, captured by a TV camera, and selected spectra: (a) $f(\mathbf{x})$, hard-limited TV image of a nut (64×64 pixels); (b) $g(\mathbf{x})$, similar image of the nut shifted (by approximately 3 pixels both horizontally and vertically) and the rotated (by 45°) surrounded by bolts and nuts of different shapes; (c) magnitude of the spectrum of the image in (a). The triangle shown in (c) identifies the part of the bispectrum defining the descriptor (refer to §3.2.2).

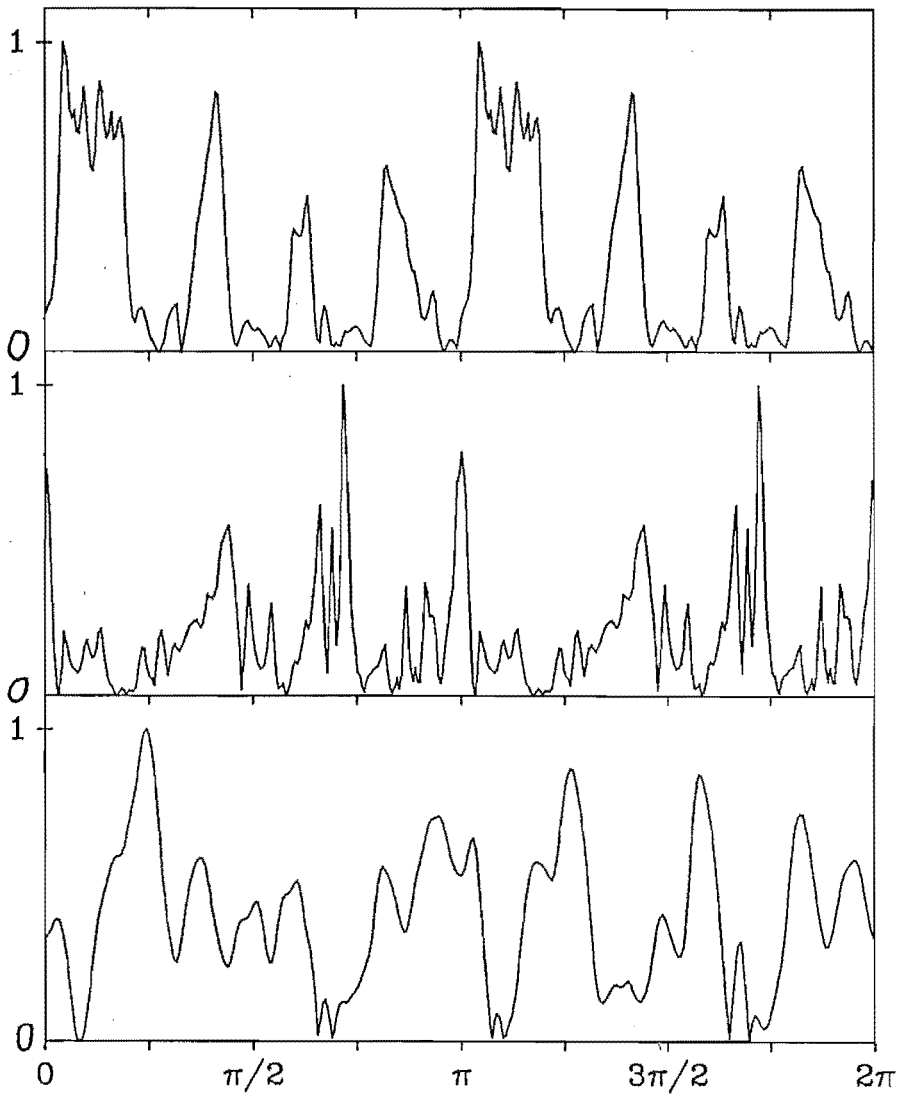


Figure 8.2. Bispectral angular correlations for the nut shown in Fig. 8.1:
Top : bispectral descriptor $F_{\rho,\tau,R}(\phi)$,
Middle : bispectral descriptor $G_{\rho,\tau,R}(\phi)$,
Bottom : angular correlation $\Gamma(F_{\rho,\tau,R}, G_{\rho,\tau,R}, \phi)$.
Note that all plotted curves represent normalized functions, i.e. the maximum of each of them is set to 1.

results obtained by employing these two descriptors is that the phase information has played an important role in the bispectral angular correlation.

Further improvement in object orientation estimation can be effected by windowing the captured TV images so as to reduce the influence of objects surrounding the dominant object (i.e.the object whose orientation is to be estimated). Computational results show that, for the nut shown in Fig. 8.1(b), a Hanning window applied to $g(\mathbf{x})$ enables the orientation of the nut to be correctly determined, i.e. $\phi = 45°$. Similar improvements have been achieved when estimating orientations of other ob-

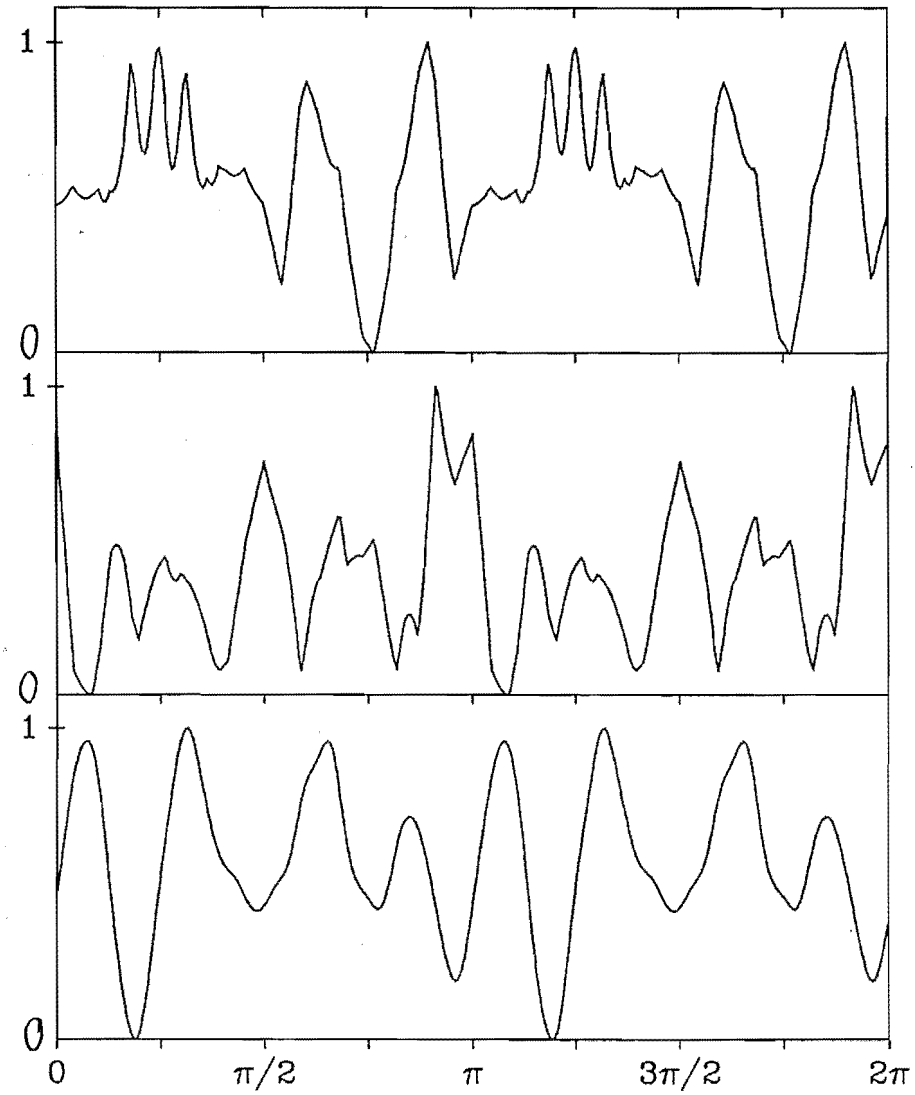


Figure 8.3. Spectral magnitude angular correlations for the nut shown in Fig. 8.1:
Top : spectral magnitude descriptor $|F_\rho(\phi)|$,
Middle : spectral magnitude descriptor $|G_\rho(\phi)|$,
Bottom : angular correlation $\Gamma(|F_\rho|, |G_\rho|, \phi)$.
Note that all plotted curves represent normalized functions, i.e. the maximum of each of them is set to 1.

jects, which suggests that windowing may be useful in practice. There will always, however, be some uncertainty over the relative positions of objects on a conveyor belt, so that one will probably only be able to employ rather weak windowing so as to avoid the possibility of distorting the image of the very object whose orientation is to be estimated.

It may be advantageous to apply the bispectral angular correlation in combination with intensity-based matching techniques of the kind described by Goshtasby [1985] for template matching employing moments (see §1.2.3). The restriction that an image should be dominated by a single object can be relaxed at the cost of increased computational effort. The zeroth order invariant moment, i.e. the mean [Goshtasby, 1985], of the template can still be used for determining possible template translations in a search area. Within a window which is positioned corresponding to each possible template translation in the search area, the angular bispectral correlation could then be invoked to match the template and to determine its rotation. The computational cost in such a matching scheme may be lower than when one relies entirely on computing central invariant moments.

Appendix A

Coronary Phantoms

In radiography, a phantom is a model of a portion of the human body which simulates the absorption of X-rays by the organs and their contents. The coronary phantom that is employed in the Cardiology Department of the Princess Margaret Hospital simulates a stenosed coronary artery filled with contrast medium. It is made of a block of Perspex which has relatively low attenuation to X-rays. Its construction is shown in Fig. A.1. In the middle of the plastic block, there is a cylindrical hole with a diameter which varies within the block. The hole is filled with the same contrast medium that is typically injected into coronary arteries, and capped at both ends.

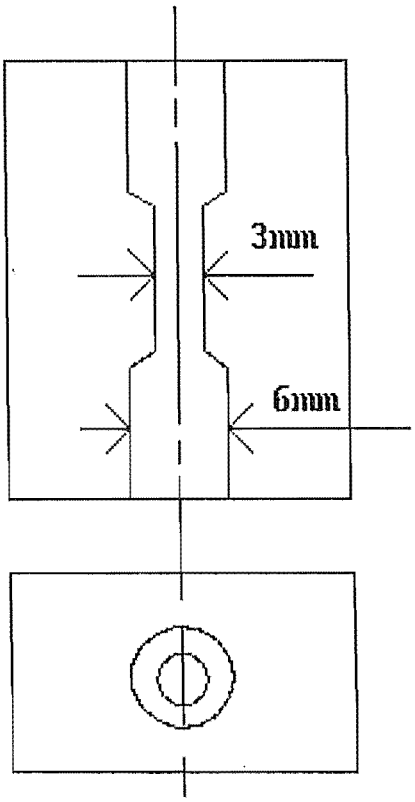


Figure A.1. The construction of a coronary phantom.

Appendix B

Removal of Electrical Interference

Australasian Physical & Engineering
Sciences in Medicine (1987) Vol. 10 No. 2

REMOVAL OF SPIKE INTERFERENCE FROM DIGITISED ANGIOGRAPHIC IMAGES

Q.X. Wu[1] and P.J. Bones[2]

1. Department of Electrical and Electronic Engineering, University of Canterbury, Christchurch, New Zealand
2. Department of Cardiology, The Princess Margaret Hospital, Christchurch, New Zealand

Abstract

We are developing a digital image processing system to augment a conventional cineangiocardiography system used in cardiac investigations. The input to our microcomputer-based system is the standard interlaced video signal used for real-time fluoroscopic display. To avoid subjecting the patient to additional radiation or radio-opaque dye infusion, data is recorded on video tape during routine cineangiographic procedures and subsequently digitised. Unfortunately significant interference to the video signal is generated by the cine motor control circuitry. The main component of the interference is a distinctive short-duration spike repeated at a rate of approximately 2 KHz. Because each spike is similar and distinct from the underlying image intensity variation, it has been possible to remove the interference very satisfactorily by software. The algorithm uses horizontal and vertical templates to detect the spikes and simple linear interpolation between video lines to remove them.

Key Words

Digital Subtraction Angiography, Digital Filtering, Cardiac image processing.

1. Introduction

The technique of digital subtraction angiography (DSA) is now well established in radiological practice, particularly for imaging stationary organs of the body (1,2,3). To perform DSA, one two-dimensional X-ray projection image is obtained with the structures of intrest opacified by radio-opaque dye (henceforth referred to as "dye"). A second image of exactly the same region, made before the dye was injected, is subtracted from the first to produce an enhanced image of the opacified structures. Early reports of the imaging technique emphasised the potential of DSA to replace arterial injections of dye with intravenous injections, thus to decrease the risk of injury to the patient and to increase the acceptability of the procedure (4). However, in practice DSA has become a complementary technique to conventional angiography and both intravenous and selective intraarterial dye injections are employed (2).

The almost constant motion of the heart makes that organ more difficult to image by DSA and the technique is not yet established in routine practice.

Generally, synchronising of data acquisition with the cardiac cycle (by, say, a trigger pulse derived from the electrocardiogram) is required (3). Several commercial DSA systems with the ability to do some cardiac imaging are available, but their high capital cost (6) makes them difficult to justify for a facility performing only cardiac investigations. We have assembled a microcomputer-based system using mainly off-the-shelf hardware items and are developing a software package to perform cardiac DSA procedures. Our work has three main aims:

1. To estimate indices of left ventricular function from central venous dye injections;
2. To assess the patency of coronary bypass grafts with a minimum of invasion (i.e. in such a way that they may be regularly assessed after surgery);
3. To estimate the relative distribution of aortic flow throughout the systemic circulation.

In the following we describe a method of removing a form of interference which we experience in the video signals received from a CGR cineangiographic system, namely frequent short-duration spikes arising in the cine motor control circuits. In section 2 we describe the hardware making up our digital image processing system. The nature of the interference is described in section 3. The methods we have used to filter out the noise are divided into "detection" and "removal" in sections 4 and 5 respectively. Finally, in section 6 we comment on the usefulness of the methods in other applications and on the efficiency of its implementation.

2. Image Acquisition

Our developmental system, which has been named the Cardiology Angiographic Processing System (henceforth referred to as CAPS), is based on the Intel Multibus I, a 16-bit microcomputer bus. The boards plugged into the bus include an Intel 80286 processor, a Matrox MIP-512 video processor and a Winchester disk controller. The microcomputer operates under Intel's RMX (for real time multitasking executive) operating system. All the major components are housed in an Intel system 310 cabinet, sited in a room adjacent to the cardiac catheterisation laboratory of The Princess Margaret Hospital, Christchurch.

Manuscript first received 27 November 1986.

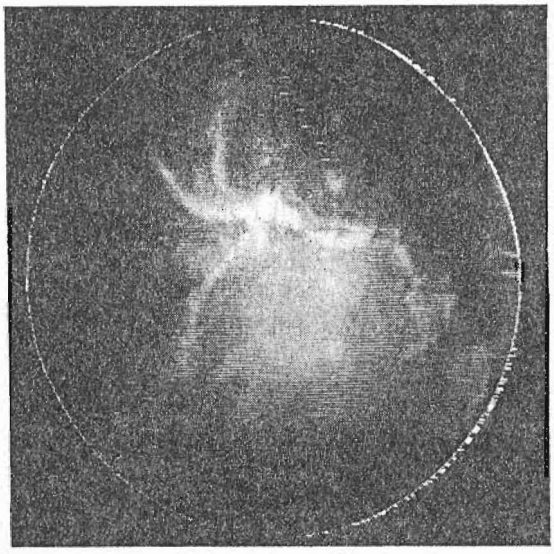


Fig.1 An image digitised during the injection of contrast medium into the left main coronary artery. The interfering spikes appear as small horizontal white or black lines within the circular intensified area.

A 75-ohm coaxial cable connects the video display chain of the CGR cineangiographic system to CAPS, supplying a standard 50Hz, 625-line interlaced video signal from the Thompson CSF vidicon camera. Selected sequences are recorded on 3/4" U-matic tape for subsequent analysis. From tape the video signal is input to the Matrox MIP-512 board which is capable of digitising video frames at a resolution of 512x512x8-bits. Each row of the digitised image corresponds to one line of the video image. Data is acquired by two methods:

1. a "scout" scan of the recorded sequence is made, monitoring the mean intensity of a small segment of the image; individual frames are chosen and then grabbed during a second replay;
2. a particular segment of interest (say 100x100 pixels) is identified and that data is acquired from each frame during the replay until a memory buffer is full.

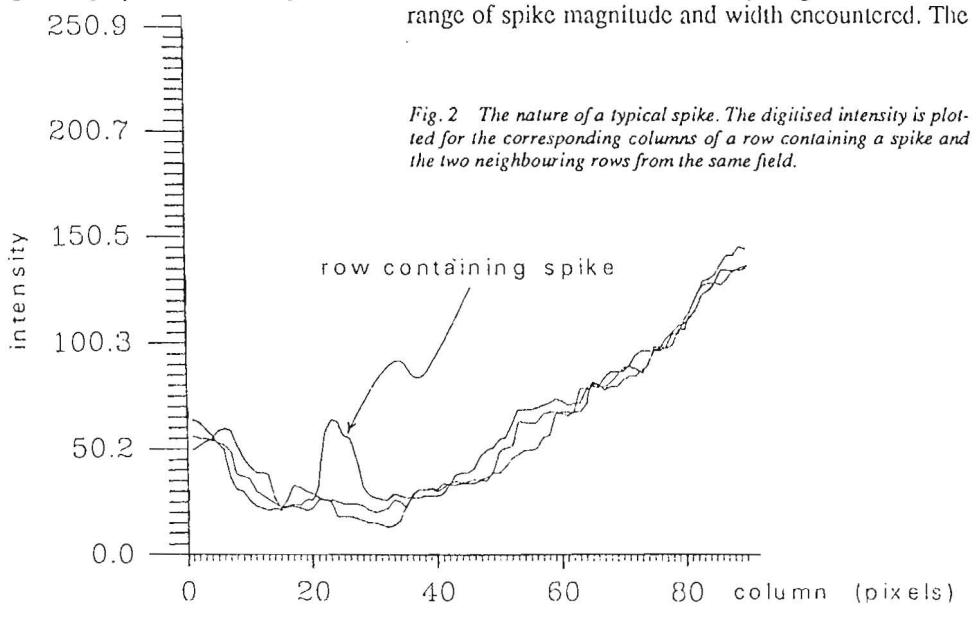
A software package locally developed for CAPS controls the Matrox board operations and manages the movement of sequences of image data on and off Winchester disk.

3. Interference

To avoid exposing the patient to additional radiation or infusing additional dye, video sequences for CAPS are recorded simultaneously with routine 35mm cine recordings. Unfortunately, the operation of the cine camera induces bursts of short-duration spikes on the video signal. The interference is not serious enough to impede the routine procedures (no interference appears on the 35mm film, only on the video signal), but it often degrades the quality of individual video frames digitised by CAPS. An example is shown in Fig. 1. The image shown was made during an injection of 75 per cent Urografin into the left main coronary artery using an 8F Cordis catheter. The spikes appear as small horizontal white or black lines (higher or lower than surrounding intensities repsectively). When one such image is subtracted from another, both sets of spikes appear in the resulting image causing considerable loss of quality.

The nature of a typical spike is illustrated in Fig. 2. The digitised intensity is plotted over a width of 90 pixels along the image row containing the spike and along the same portion of the neighbouring rows from the same video interlaced field (remember that an interlaced video frame is made up of two fields - see, for example, (2)). The spike is clearly distinct from underlying image features and random noise appearing on the video signal. The spikes appear in bursts with a repetition rate of approximately 2 KHz (corresponding to one spike every 14 rows). While the spikes are not all of the same shape and height they do show considerable similarity. Fig. 3 shows the range of spike magnitude and width encountered. The

Fig. 2 The nature of a typical spike. The digitised intensity is plotted for the corresponding columns of a row containing a spike and the two neighbouring rows from the same field.



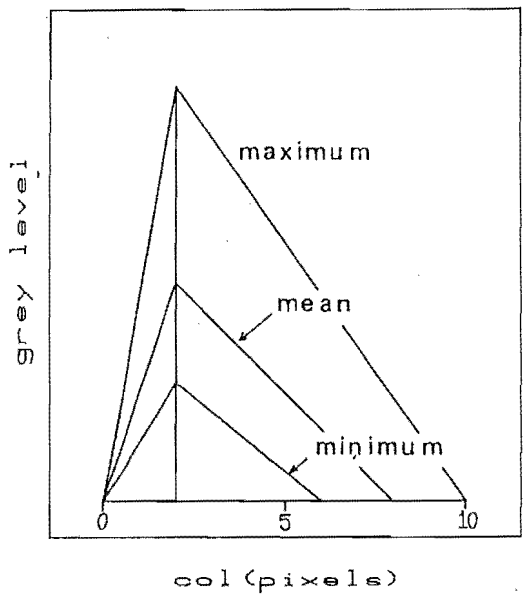


Fig. 3 The range of spike magnitude and width encountered (relative magnitude only).

width varies from 600 to 1000 nsec (one horizontal pixel corresponds to 100 nsec) and the maximum amplitude is 3 times the minimum

4. Spike Detection Algorithm

4.1 Basic Theory

The detection of algorithm is based on the "matched filter" concept (7). Given a function f(x) which is both finite and real, we can write

$$ff(x)=\int_{-\infty}^{\infty} f(t)f(t+x)dt \qquad (1)$$

where ff(x) is the autocorrelation of f(x). The autocorrelation theorem (8) states that

$$|\ F(u)\ | \ <-\ ->ff(x) \qquad (2)$$

where <- -> interconnects the two members of a one-dimensional Fourier transform pair, and it follows from the definition of ff(x) (8) that

$$ff(-x)=ff(x) \qquad (3)$$

and

$$ff(0)=\int_{-\infty}^{\infty} f(t)^2\,dt \qquad (4)$$

Here Eq. (1) and (4) also infer that ff(0) is the maximum of ff(x) and is in the centre of the extent of ff(x).

These properties of the autocorrelation function suggest a method of designing a matched filter. If we know the characteristics of a certain additive noise (i.e. its shape and extent) which appears in the signal, we can design a template with the same shape and extent as the noise. Supposing the noise (spike) has a shape as Fig. 4(a), the template would be as drawn in Fig. 4(b). The mean amplitude of the template is made

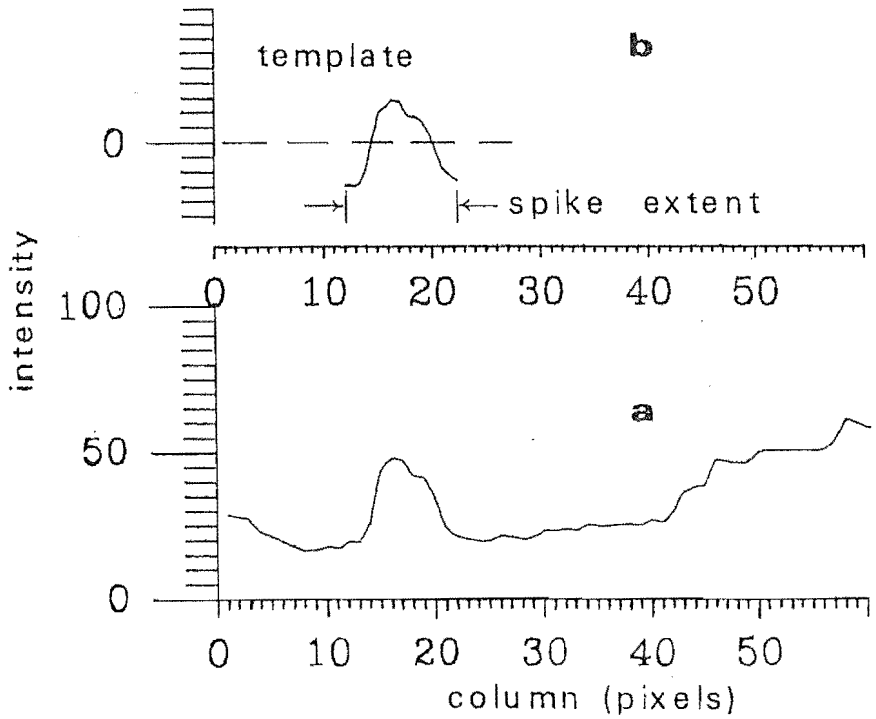


Fig. 4 Designing a matched filter. To detect the spike shown in a), the template is constructed as shown in b).

zero by altering the DC level and the peak-to-peak amplitude chosen for convenience in computation. The template is correlated with the signal by

$$C_{fg}(x) = \int_b^a g(t)f(t+x)dt \qquad (5)$$

where g(x) is the template which extends from a to b and $C_{fg}(x)$ is the cross-correlation of g(x) with f(x). Whenever the cross-correlation process encounters a spike, $C_{fg}(x)$ is locally maximum.

4.2 Practical Implementation

To apply the matched filter derived above to the digitised image data, we first write Eq. (5) in discrete form:

$$C_{fg}(i) = \sum_{k=-K}^{K} g(k)f(k+i) \qquad (6)$$

and express the template g(k) in the form $[g_k, g_{k-1}, \ldots g_0, \ldots g_{-k+1}, g_{-k}]$. In Eq. (6) both indices i and k relate to successive pixel positions along a row of the image. In section 3, the spikes were described as having an extent of up to 10 pixels, suggesting a value of 5 for K. To perform Eq. (6) over the entire image with K=5 would require several million arithmetic operations and be necessarily slow. To speed up execution, we simplified the template to [-1,0,0,2,0,0,-1], i.e. only 3 non-zero coefficients, and thereby reduced the number of operations to approximately one million per 512x512 image. Two examples of the operation of the matched filter are shown in Fig. 5. In each part of the Fig. 5, the image intensity (lower) and the cross-correlation of the intensity with the template are shown for a portion of a row containing a spike. The cross-correlation can be seen to peak at the time of occurrence of the spike.

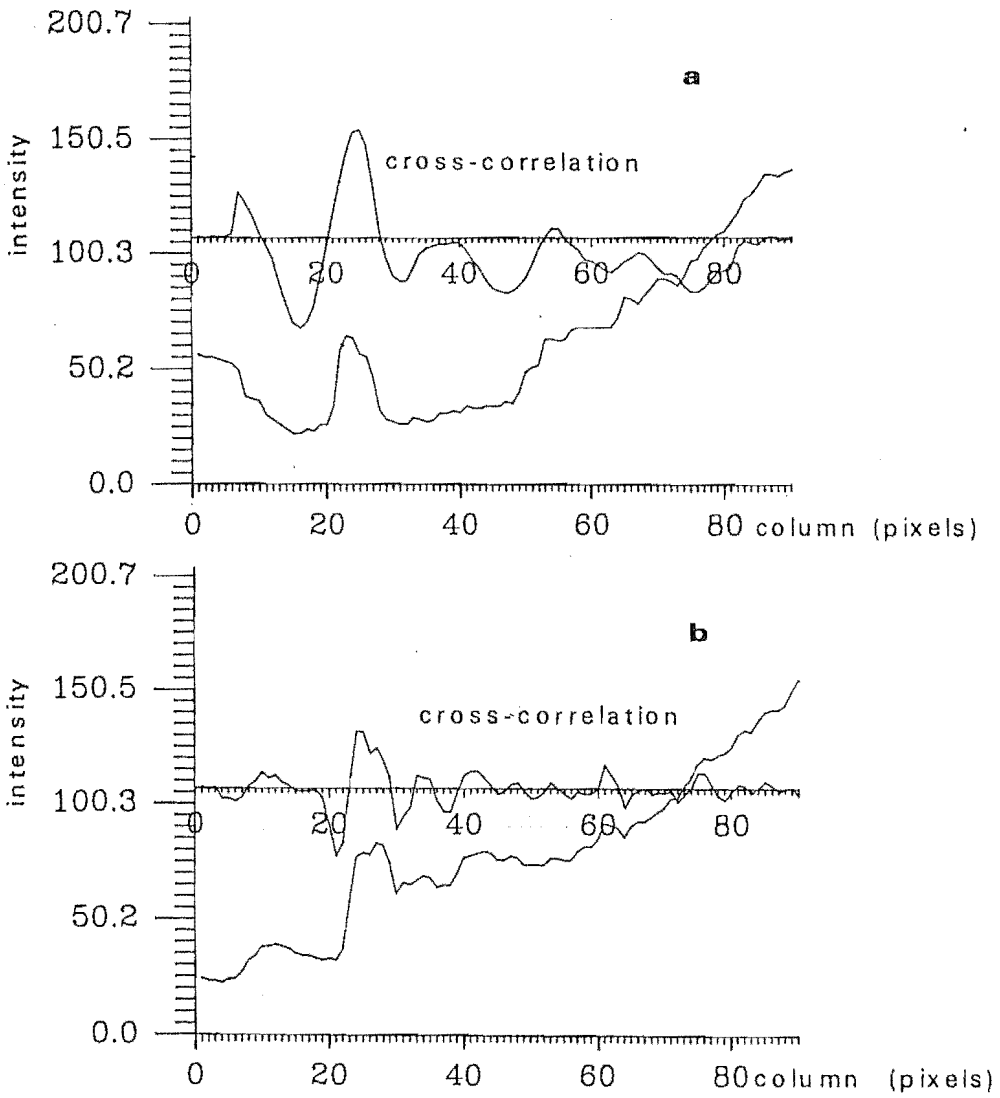


Fig. 5 Two examples of the operation of the matched filter. In a) and in b) the image intensity (lower trace) and the cross-correlation of the intensity with the template are shown for a portion of a row containing a spike.

In our algorithm, the detection of a spike is considered likely if the cross-correlation exceeds a threshold value. The detection is confirmed by comparing the pixel concerned P(x,y) with the corresponding pixels in the previous and the next rows in the field, P(x,y,-2) and P(x,y+2) respectively, where x and y are integers expressing position in terms of columns and rows. The criteria for confirmation are:

$$\begin{array}{lll} |\ P(x,y) - P(x,y\text{-}2) & |> & \text{threshold} \\ |\ P(x,y) - P(x,y\text{+}2) & |> & \text{threshold} \\ |\ P(x,y\text{+}2) - P(x,y\text{-}2) & |< & \text{threshold} \end{array} \qquad (7)$$

where threshold takes the same value as for the matched filter.

5. Spike Removal Algorithm

Normally more then one pixel within a spike may meet the detection criteria described above. The first step in the removal algorithm is therefore to locate the approximate centre of the spike. This is taken to be the median x coordinate of the set of detected pixels in the row, using the *a priori* expectation that at most one spike will appear in a row.

The method we have adopted to remove spikes after detection is to replace the pixels in the same row and within the extent of spike with values linearly interpolated from the neighbouring rows. Thus if a spike is detected at (x,y),

$$P(x', y) <\text{- - -}\ (\ P(x', y+2) + P(x', y-2)\ )/2 \qquad (8)$$

for all x' given by $2|x'-x| \leq R$, where R is the removal width in pixels. Note that the interpolation uses the neighboring rows from the same video field so that the interpolation is not influenced by any fluctuation in mean intensity between the interlaced video fields.

Fig. 6 shows the identical image to that in Fig. 1 after the spike detection and removal program has been executed. The program has proved very successful at removing the spikes. The original implementation, written in PASCAL with low-level PL/M procedures for accessing the Matrox MIP-512, takes approximately 15 seconds to process the complete 512x512 image. A version written entirely in PL/M and making better use of the 80286 processor's instruction set has been completed and this reduces the execution time to approximately 10 seconds.

6. Conclusion

The matched filter method has proved very successful at detecting the cine motor generated spikes in the digitised video data. The first stage of the algorithm, i.e. correlating the template along each row of the image, produces a significant number of false positives, but these are virtually all eliminated by the application of the criteria in Eq.(7). Thus the matched filter method has been considerably enhanced by the application of *a priori* knowledge of the expected repetition rate of the spikes. This idea could be extended to other one- or two-dimensional filtering applications by the use of a variable threshold function based on the expectation of when or where the next event is to occur.

Ideally, any form of interference is best removed at the source ("prevention is better than cure"). In the case of CAPS however, we were determined to work with the images obtainable without modification to the commercial cineangiographic system and without interfering with the routine protocol of the investigation. The success of the spike detection and removal

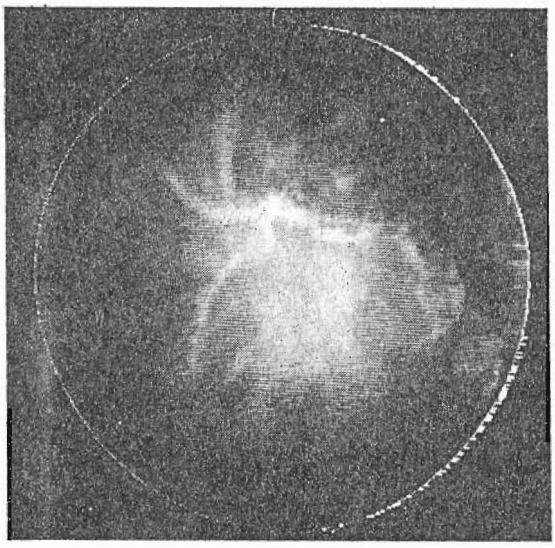


Fig. 6 The identical image to that in Fig. 1 after execution of the spike detection and removal program.

algorithm has meant that our DSA work has not suffered as a result of the interference. The algorithm cannot be implemented in real time with the existing CAPS hardware. However the addition of a further MIP board to the system would allow convolution operations to be performed fast provided sufficient storage for at least 3 images is available (9). The matched filter part of the detection algorithm is effectively a convolution with a 3-element template. This and other operations involved in the complete algorithm could be implemented in less than 15 refresh periods, i.e. less than one second.

7. Acknowledgements

One of the authors, P.J. Bones, is supported by a Senior Fellowship from the National Heart Foundation of New Zealand.

References

1. Kruger R.A., Liu P.Y., *Digital angiography using a matched filter.* IEEE Trans on Med. Imag. (1982), MI-1, 16.

2. Levin D.C., Schapiro R.M., Dunham L, Harrington D.P. and Ergun D.L., *Digital Subtraction Angiography: principles and pitfall of image improvement techniques.* Amer. J. Roentgenology (1984), 143, 447.

3. Meaney T.F., et al. *D.S.A. of the human cardiovascular system.* Amer. J. Roentgenology (1980), 135, 1153.

4. Riederer J.J., Kruger R.A., *Intravenous digital subtraction: a summary of recent developments.* Radiology (1983), 147, 663.

5. Lipton M.J., Gould R.G., Herfken R.J., *Advances in and future directions of digital subtraction angiography. In* "Contrast Media in Digital Radiology", Felix R. (Ed), Netherlands, Excerpta Medica (1983), p51.

6. Evens R.G., *Cost of digital angiography in 1983: Impact on planning and charges. In* "Contrast Media in digital Radiography", Felix R. (Ed), Netherlands, Excerpta Medica (1983), p396.

7. Lynn P.A., *"An Introduction to the Analysis and Processing of Signals"*, Sec. Ed., MACMILAN, Hong Kong (1982).

8. Bates R.H.T., and McDonnell M.J., *"Image Restoration and Reconstruction"* Clarendon Press, Oxford, (1986), chapter 2.

9. *Matrox OEM boards and systems date catalogue.* Montreal, Matrox Electronics systems Ltd (1985), section 1, p73.

Appendix C

Video-Rate Processing of Image Sequences

VIDEO-RATE PROCESSING WITH CONVENTIONAL IMAGE ACQUISITION HARDWARE

P.J Bones
Q.X. Wu
R.G. Lane

Department of Electrical & Electronic Engineering, University of Canterbury.

ABSTRACT

A number of plug-in image acquisition/display boards are available for the major microcomputer bus families with the least expensive PC-bus board costing in the vicinity of NZ$3,000. In processing sequences of images at the video rate of 25 frames/second, major restrictions are the rate at which data can be transferred across the host system's bus and the sheer volume of data concerned.

This paper reports success with two methods of sequence processing (often used in combination) using only conventional image acquisition hardware. The first is real-time processing of one or more areas-of-interest within each frame. Applications for the method include monitoring intensity as a function of time and detection of motion. The second method is the assembly of a sequence of consecutive frames by means of multiple replays of a video tape. The latter has been successful in analysing X-ray images of the beating heart. A novel method for reanimating the digitised sequence has also been established.

It is thus shown that useful tasks can be performed on video-rate image sequences, without resorting to high-cost imaging engines or parallel architectures.

1. INTRODUCTION

Many applications of image processing require observing phenomena in real-time, or at least "video-rate" processing, a term coined to encompass sequences of images replayed from video tape. Examples are: i) detecting the arrival of an object or objects within a camera's field of view; ii) monitoring the concentration of a substance (e.g. a pollutant) within a flowing fluid channel; iii) tracking the motion of a simple object over a small range; and iv) monitoring the orientation, shape or size variation of a simple object.

Several manufacturers produce relatively inexpensive plug-in image acquisition/display boards for the major microcomputer bus families, e.g. IBM PC, VME, Multibus, Qbus, etc. The boards all feature at least 512 x 512 x 8-bit digitising resolution, the ability to "grab" images from a standard video signal in real-time into one or more frame stores, and look-up tables to transform incoming and outgoing data. Often other features like zoom and pan are incorporated.

Most software support packages available for these boards are oriented towards the acquisition and processing of single frames; real-time processing is not normally incorporated. The reason is clearly that there is a incompatibility between the (typically) 10 MHz rate of acquiring, storing and recirculating digital samples in the board's frame store, and the rate at which data can be moved elsewhere in the microcomputer system via the bus [1, 2]. The sustainable bus transfer rate is of the order of 20 times slower than the acquisition rate. Secondly, the microcomputer host typically has an amount of available RAM comparable to the size of the frame store (and therefore can accommodate only a single frame at a time). Winchester disk (or any other form of mass storage currently found on microcomputer systems) has an even slower rate of sustainable data transfer. While the capability of microcomputer hardware is continually improving, continuous acquisition and processing of full-frame video data on a microcomputer system is some way off.

Despite the limitations discussed above, the conventional plug-in image acquisition boards can perform useful video-rate tasks. This paper describes two methods employed by the authors in a variety of applications. Examples discussed have all been implemented on the Cardiology Angiographic Processing System (CAPS) developed by the authors at The Princess Margaret Hospital, Christchurch [3]. CAPS is based on an Intel System 310 with a Multibus I backplane, 80286/80287 processor and Matrox MIP-512 image acquisition/display boards.

2. AOI METHOD

Many of the tasks which may be required of a machine vision system require, at any one time, only data from a small region within the image to be processed. For these tasks one or more rectangular areas-of-interest (AOI) can be identified within the image. During processing, only the data from within the AOI(s) is transferred to the host's RAM, with the transfer being initiated by an interrupt coincident with the commencement of the vertical blanking period. Boards in which the frame stores are mapped onto the host microprocessor's memory address space allow the use of efficient block move instructions (such as MOVS in the Intel 8086 family of processors) to shift a set of contiguous pixels in an image row from the frame store to the host's RAM (see Fig. 1). This clearly

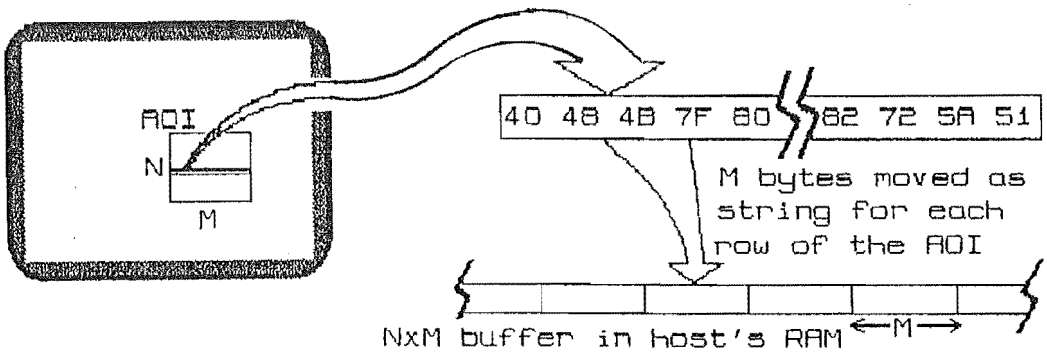


Figure 1. Data within an AOI (N rows by M columns) is moved as N row-wise operations to a RAM buffer.

indicates that row-wise operations are in general faster than column-wise operations. For example, to execute an efficient procedure to transfer a complete row of 512 bytes on CAPS takes 1.4 msec, wheras the equivalent column-wise operation takes 43 msec. The same system can transfer a 100 x 100 pixel AOI in 33 msec, performing 100 row-wise operations.

Two examples illustrate applications of the AOI method.
Example 1. During a pre-recorded video sequence of an X-ray examination of the left ventricle of the heart, it is necessary to monitor the mean intensity within a small AOI positioned appropriately. For this simple task, there is no need to transfer the pixel data to the host's RAM. At each frame interrupt, the AOI is scanned row-by-row and the mean of the pixel intensities is calculated. To detect if the X-ray system has been activated, the mean is compared with a preset threshold; to record the intensity fluctuations due to the beating of the heart, the mean for each frame is stored in an array.
Example 2. A camera is positioned vertically over a conveyor belt so that the direction of conveyance is from the top towards the bottom. The "AOI" is chosen to be a single row of the image near the centre. This row is transferred to the host's RAM each frame for further processing. The video camera is thereby simulating a line-scan camera, with a line being acquired every 40 msec. Provided the lighting system is suitable and sufficient contrast exists between the conveyor surface and the objects being transported, such a system might (in real time) reasonably detect the arrival of objects and discriminate coarse differences in the shape and orientation of objects [1].

3. MULTIPLE-REPLAY METHOD

While for many industrial applications of image processing, real-time response is required, many scientific and medical applications tolerate a degree of post-processing [see, for example, Refs. 3, 4]. Therefore storing video sequences on video tape for later analysis is feasible. In general it is possible for some discrete event to be recorded on the tape, preceding the sequence to be analysed, which may be used to trigger the acquisition process. Examples of suitable trigger events are a step-like change in intensity at some fixed image location or a control pulse encoded onto the accompanying audio track.

The multiple-replay method is best illustrated with an example (see Fig. 2). Suppose that a short sequence of images, each being a 256 x 256 pixel area from the grabbed frame, is to be digitised and stored in a disk file. 256 KB is available in the host's RAM and a 1 MB-sized file is open for sequential access. The following steps are performed:

1. Rewind the tape and play from before the trigger event.
2. After detection of the event wait a preset interval.
3. Grab every 4th frame and transfer the required 256 x 256 area of data (64 KB) to the host's RAM buffer until the buffer is full.

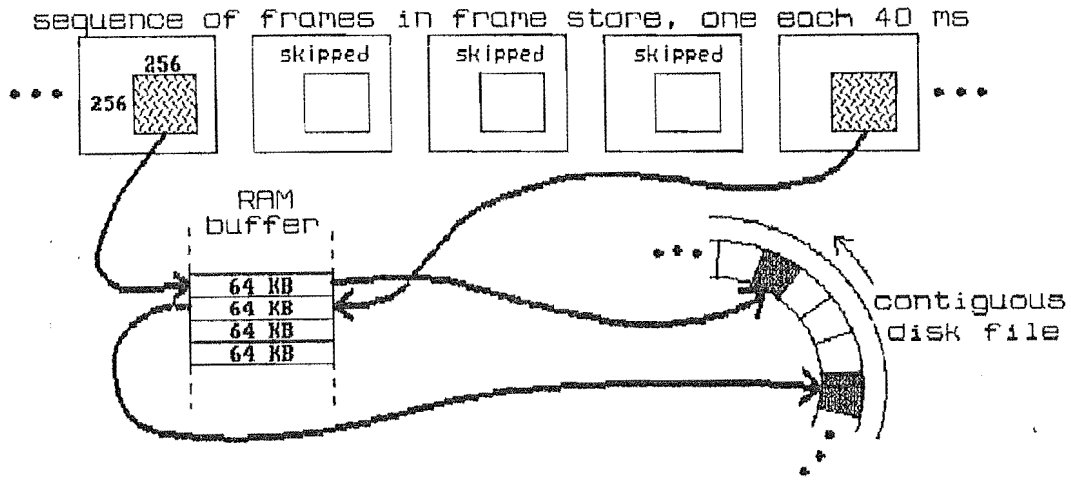


Figure 2. Illustration of multiple-replay method of acquiring a sequence of large image areas and assembling in a disk file.

4. Write the first 64 KB of buffered data to the disk file, skip 3 x 64 KB in the file, write the second 64 KB of buffered data, etc.
5. Repeat steps 1 to 4 a further 3 times with (for each replay) a delay of one additional frame in the grabbing sequence plus a further 64 KB offset in the disk file.

At completion of the process, a sequence of 16 consecutive frames of the 256 x 256 pixel image has been assembled on disk.

An example of the use of the multiple-replay method is the analysis of the heart's pumping performance developed by one of the authors (P.J.B.). A video signal is recorded from an intensified X-ray system immediately following the rapid injection of radio-opaque dye into the left-ventricular chamber of the heart. Of interest is the relative change of volume (as estimated from the 2-dimensional projection of the dye-filled interior onto the imaging plane) during the ejection period of a typical beat. Since this period is 0.5 second or less, a sequence of 12 consecutive frames is sufficient to capture the complete ejection period. A total of 5 replays of the video tape are used: the first to establish the precise time from trigger point to the beat to be processed, 3 to assemble the sequence of 12 consecutive frames and the last to capture 4 further frames to be available for use as masks for subtraction (used to improve contrast at the chamber boundary). Fig. 3 shows a typical sequence.

A novel method has been developed for displaying the image sequence thus obtained. All 12 frames are "dealt" (i.e. like a pack of cards) into the image frame store (see Fig. 3). Then each frame is displayed (using hardware zoom and pan) in the centre of the display monitor in sequence, with the switch from one frame to the next being made during the blanking period of the video. Further, the sequence of 12 is alternately "played" forward and backward to create the illusion of a continuously beating heart. The outline of the chamber is thereby observed for subsequent analysis of the pumping performance.

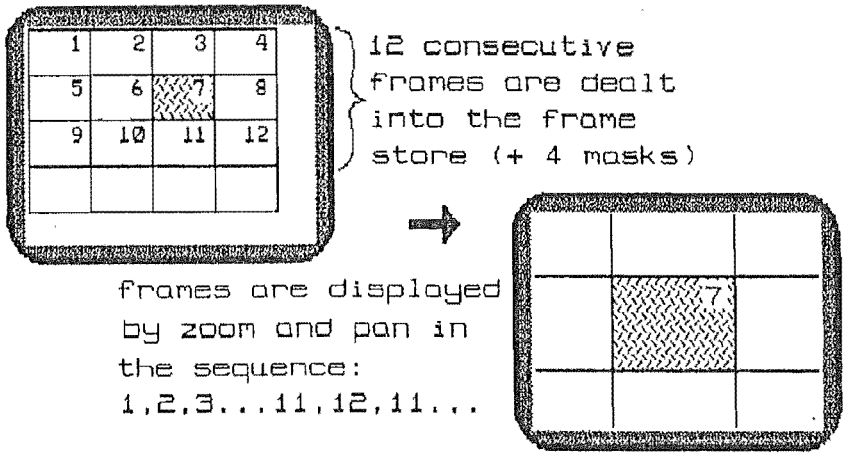


Figure 3. Display of beating heart: sequence of X-ray images is "dealt" into the frame store then animated with zoom and pan.

4. DISCUSSION

The drive towards better and faster facilities for processing images is to the fore in promoting the development of multi-processing hardware (and software) [5]. While many developments are directed towards a wide range of applications, some systems come into the category of "imaging engines" with very large frame stores and multiple buses [see, for example, Refs. 4, 6]. While many of these exotic machines do find a market, they are necessarily extremely costly. Their complexity also generally means that program development is slow (compilers for high-level languages may be unavailable or difficult to use). With conventional image acquisition plug-in boards, a handful of procedures - supplied with most boards - for carrying out low level tasks (e.g. to transfer a block of pixels from frame store to RAM) are all that is required for testing an algorithm. Only when this approach proves to be too slow for an application, does the more special-purpose hardware become cost effective.

For any method of image sequence acquisition to be successful, it is necessary for the video data to be of adequate quality. Since the only conceivable reason for wanting to acquire sequences is to observe intensity fluctuations (most commonly associated with motion), the video camera must clearly be able to faithfully reproduce these changes. A common problem is ghosting caused by appreciable persistence in vidicon cameras. For a discussion of the properties of current video camera types, see [1, 2]. Observations of moving objects must take into account that alternate rows of the image are from alternate interlace fields; thus the **i**th row was recorded 20 msec earlier (or later) than the **(i + 1)**th. The two fields of a frame may be considered as two distinct images, or one field may be discarded.

The multiple-replay method, or any method acquiring volumes of image data greater than the available RAM, requires efficient transfer of data to and from disk file. The CAPS software uses 256 KB contiguous files (i.e. sufficient for one 512 x 512 x 8-bit image) for storage of all sequences to minimise overhead in disk transfers (sequences may span several files). The files are never deleted (just overwritten) to ensure that they maintain their contiguity. Transferring a sequence of 25 frames, each 100 x 100 in size, from disk to frame store takes approximately 3 seconds, while the equivalent operation for a single full-frame image takes 2 seconds.

5. CONCLUSIONS

Many applications requring real-time response have quite modest requirements in terms of the area of image actually required for processing. Other applications, while needing a substantial amount of data from each video frame, can be processed offline; we coin the term "video-rate" for this type of processing. The two methods of image acquisition described here may be implemented on virtually any of the (relatively inexpensive) plug-ins offered for the major microcomputer buses. They have been found to be successful in a number of applications associated with the processing of X-ray images of the heart.

The authors acknowledge the generous support of the National Heart Foundation of New Zealand for the work described. P.J. Bones was until recently a Senior Fellow of the Foundation, Q.X. Wu is receiving their support for his postgraduate study.

REFERENCES

[1] B.G. Batchelor, D.A. Hill, D.C. Hodgson, "Automated Visual Inspection", U.K., I.F.S. (Publications) Ltd, 1985.

[2] P. Dunbar, "Machine vision", BYTE, vol. 11(1), Jan. 1986, pp 161 - 173.

[3] Q.X. Wu, P.J. Bones, "Removal of spike interference from digitised angiographic images", Australasian Physical & Engineering Sciences in Medicine, vol. 10, 1987, pp 96 - 101.

[4] K.H. Hohne, U. Obermoller, M. Riemer, G. Witte, "Fourier domain techniques for digital angiography of the heart", IEEE Trans. on Medical Imaging, vol. MI-3(2), June 1984, pp 62 - 67.

[5] K-S. Fu, "VLSI for Pattern Recognition and Image Processing", Heidelberg, Springer-Verlag, 1984.

[6] C. Bowman, "A Kiwi with speed and vision", New Electronics (A.G.M., Auckland), March 1988, pp 41 - 43.

References

Adiv, G. (1985), 'Determining three-dimensional motion and structure from Optical Flow generated by several moving objects', *IEEE Trans. Pattern Anal. and Machine Intell.*, Vol. PAMI-7, No. 4, Pp. 384–401.

Aggarwal, J.K. (1984), 'Motion and time-varying imagery', *Computer Graphics*, Junuary, Pp. 20–21.

Aggarwal, J.K. and Duda, R.O. (1975), 'Computer analysis of moving polygonal images', *IEEE Trans. Computers*, Vol. C-24, No. 10, Pp. 966–976.

Aggarwal, J.K. and Martin, W.N. (1983), 'Dynamic scenes analysis', In Huang, T. (Ed.), *Image Sequence Processing and Dynamic Scene Analysis*, Springer-Verlag, Berlin, Pp. 40–74.

Aggarwal, J.K., Davis, L.S. and Martin, W.N. (1981), 'Correspondence processes in dynamic scene analysis', *Proc. IEEE*, Vol. 69, No. 5, Pp. 562–572.

Agin, G.J. (1979), 'Knowledge-based detection and classification of vehicles and other objects in aerial road images', In Baumann, L.S. (Ed.), *Proc. Image Understanding Workshop*, Science Applications, Inc., Arlington/VA 22209, April, Pp. 66–71.

Akaike, H. (1966), 'Note on higher order spectra', *Annuals of the Institute of Statistical Mathematics*, Vol. 18, Pp. 123–126.

Albert, A. (1972), *Regression and the Moore-Penrose pseudoinverse*, Academic, New York.

Alliney, S. and Morandi, C. (1986), 'Digital image registration using projections', *IEEE Trans. Pattern Anal. and Machine Intell.*, Vol. PAMI-8, No. 2, Pp. 222–233.

Anderson, R.C. and Anderson, C.S. (1986), 'Signal processing using only fourier phase', *Optical Engineering*, Vol. 25, Pp. 1316–1319.

Anderson, C.S. and Anderson, R.C. (1987), 'Comparison of phase-only and classical matched filter scale sensitivity', *Optical Engineering*, Vol. 26, No. 3, Pp. 276–279.

Ballard, D.H. (1981), 'Generalizing the Hough transform to detect arbitrary shapes', *Pattern Recognition*, Vol. 13, No. 2, Pp. 111–122.

Ballard, D.H. and Brown, C.M. (1982), *Computer Vision*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Balter, S., Ergun, D., Tscholl, E., Buchmann, F. and Verhoeven, L. (1984), 'Digital subtraction angiography: fundamental noise characteristics', *Radiology*, Vol. 152, No. 1, Pp. 195–198.

Barnea, D. and Silverman, H.F. (1972), 'A class of algorithms for fast digital image registration', *IEEE Trans. Computers*, Vol. C-21, No. 2, Pp. 179–186.

Barrow, H.G. and Tenenbaum, J.M. (1981), 'Computational vision', *Proc. IEEE*, Vol. 69, No. 5, Pp. 572–595.

Bartelt, H., Lohmann, A.W. and Wirnitzer, B. (1984), 'Phase and amplitude recovery from bispectra', *Applied Optics*, Vol. 23, No. 18, Pp. 3121–3129.

Batchelor, B.D., Hill, D.A. and Hodgson, D.C. (1985), *Automated Visual Inspection*, IFS(Publication) Ltd., UK.

Bates, R.H.T. (1982), 'Astronomical speckle imaging', *Physics Reports*, Vol. 90, No. 3, Pp. 203–297.

Bates, R.H.T. (1987), 'Aspects of the Erlangen-bispectrum', *Optik*, Vol. 76, No. 1, Pp. 23–26.

Bates, R.H.T. and Fright, W.R. (1985), 'Tow-dimensional phase restoration', In Price, J.F. (Ed.), *Fourier Techniques and Applications*, Plenum Publishing Corp., Pp. 121–148.

Bates, R.H.T. and McDonnell, M.J. (1986), *Image Restoration and Reconstruction*, Clarendon Press, Oxford, London.

Bates, R.H.T., Gough, P.T. and Napier, P.J. (1973), 'Speckle interferometry gives holograms of multiple star systems', *Astronomy and Astrophysics*, Vol. 22, No. 2, Pp. 319–320.

Bates, R.H.T., Garden, K.L. and Peters, T.M. (1983), 'An overview of computerized tomography with emphasis on future developments', *Proc. IEEE*, Vol. 71, March, Pp. 256–372.

Bhanu, B. (1987), 'CAD-Based robot vision', *Computer*, Vol. 20, August, Pp. 13–16.

Bhanu, B. and Ho, C. (1987), 'CAD-Based 3d object representation for robot vision', *Computer*, Vol. 20, August, Pp. 19–35.

Binford, T.O. (1982), 'Survey of model-based image analysis systems', *The International Journal of Robotics Research*, Vol. 1, No. 1, Pp. 18–64.

Bjorkholm, P., Annis, M., Frederick, E., Stein, J. and Swift, R. (1981), 'Digital radiography: spatial and contrast resolution', In *Application of Optical Instrumentation in Medicine IX, Vol 273*, SPIE, Pp. 103–107.

Bogert, B.P., Healy, M.J.R. and Tukey, J.W. (1963), 'The quefrency analysis of time series for echoes: cepstrum, pseudo-autocovariance, cross-cepstrum, and saphe cracking', In Rosenblatt, M. (Ed.), *Proc. Symp. Time Series Analysis*, John Wiley & Sons, Inc., New York, Pp. 209–243.

Bogren, H.G. and Bursch, J.M. (1984), 'Digital angiography in the diagnosis of congenital heart disease', *Cardiovasc. Intervent. Radiol*, Vol. 7, Pp. 180–188.

Bones, P.J. (1981), *Contributions to electrocardiographic science*, PhD thesis, University of Canterbury, Christchurch, New Zealand.

Bones, P.J., Wu, Q.X. and Lane, R.G. (1988), 'Video-rate processing with conventional image acquisition hardware', In *Proceedings of the National Electronics Conference of New Zealand (NELCON '88)*, Christchurch, New Zealand.

Bracewell, R.N. (1978), *The Fourier Transform and its Applications*, McGRAW-HILL, 2 ed.

Brennecke, R., Hohne, H.J., Moldenhauer, K. and Heintzen, J.H. (1978), 'Improved digital real time processing and storage techniques with applications to intravenous contrast angiography', In *Computer in Cardiology, IEEE Proceeding*, Pp. 191–194.

Brillinger, D.R. (1965), *Time Series Analysis*, Holt. Rinehardt and Winston, New York.

Brody, W.R. (1981), 'Hybrid subtraction for improved arteriography', *Radiology*, Vol. 141, No. 3, Pp. 828–831.

Brody, W.R., Cassel, D.M. and Sommer, F.G. (1981), 'Dual energy projection radiography: initial clinicle experience', *Amer. Jour. of Roentgenology*, Vol. 137, No. 2, Pp. 201–205.

Buonocore, E., Meaney, T.F., Borkowski, G.P., Pavlicek, W. and Gallagher, J. (1981), 'Digital subtraction angiography of the abdominal aorta and renal arteries. comparison with conventinal aortography', *Radiology*, Vol. 139, No. 2, Pp. 281–286.

Burrus, C.S. and Parks, T. (1984), *DFT/FFT and Convolution Algorithms - theory and implementation*, John Wiley & Sons, New York.

Bursch, J.H., Hahne, H.J., Brennecke, R., Gronemeier, D. and Heintzen, P.H. (1981), 'Assessment of arterial blood flow measurements by digital angiography', *Radiology*, Vol. 141, No. 1, Pp. 39–47.

Cafforio, C. and Rocca, F. (1979), 'Tracking moving objects in TV images', *Signal Proc.*, Vol. 1, Pp. 133–140.

Cafforio, C. and Rocca, F. (1983), 'The differential method for image motion estimation', In Huang, T.S. (Ed.), *Image Sequence Processing and Dynamic Scene Analysis*, Springer-Verlag, Berlin, Pp. 104–124.

Casasent, D. (1981), 'Pattern recognition: a review', *IEEE spectrum*, Vol. 18, March, Pp. 28–33.

Casasent, D. (1985), 'Coherent optical pattern recognition: a review', *Optical Engineering*, Vol. 24, No. 1, Pp. 26–32.

Casasent, D. and Psaltis, D. (1977), 'New optical transformations for pattern recognition', *Proc. IEEE*, Vol. 65, No. 1, Pp. 77–84.

Castro, E.D., Cristini, G., Martelli, A., Morandi, C. and Vascotto, M. (1987), 'Compensation of random eye motion in television opthalmoscopy: preliminary results', *IEEE Trans. Medical Imaging*, Vol. MI-6, Pp. 74–81.

Chellappa, R. and Sawchuk, A.A. (Eds.) (1985), *Digital Image Processing and Analysis: Volume 2: Digital Image Analysis*, IEEE Computer Society.

Childers, D.G. (Ed.) (1978), *Modern Spectrum Analysis*, IEEE Press.

Cochran, W.T., Cooley, J.W., Favin, D.L., Helms, H.D., Kaenel, R.A., Maling, W.W.L.G.C., Nelson, D.E., Rader, C.M. and Welch, P.D. (1967), 'What is the Fast Fourier Transform?', *Proc. IEEE*, Vol. 55, No. 10, Pp. 1664–1674.

Collins, S.M. and Skorton, D.J. (1986), *Image Processing Systems*, McGraw-Hill, New York, Chap. 7, Pp. 121–132.

Cooley, J.W., Lewis, P.A.W. and Welch, P.D. (1967), 'Historical notes on the Fast Fourier Transform', *Proc. IEEE*, Vol. 55, No. 10, Pp. 1675–1677.

Cornsweet, T.N. (1971), *Visual Perception*, Academic Press, New York.

Cottrell, D.M., Davis, J.A., Schamschula, M.P. and Lilly, R.A. (1987), 'Multiplexing capabilities of binary phase-only filter', *Applied Optics*, Vol. 26, Pp. 934–937.

Dainty, J.C. and Northcott, M.J. (1986), 'Imaging randomly translating object at low light levels using the triple correlation', *Optics Communications.*, Vol. 58, Pp. 11–14.

Davis, L.S. and Rosenfeld, A. (1976), 'An application of relaxation labelling to spring-loaded template matching', In *Proc. Internat. Joint Conf. Pattern Recognition 3rd*, Pp. 591–597.

Davis, J.A., Heissenberger, G.M., Lilly, R.A., Cottrell, D.M. and Brownell, M.F. (1987), 'High efficiency optical reconstruction of binary phase-only filters using the Hughes liquid crystal light value', *Applied Optics*, Vol. 26, No. 5, Pp. 929–933.

Davison, M. and *et. al* (1977), *The Physics of Radiodiagnosis*, Scientific Report Series–6, The Hospital Physicists' Association, 47 Belgrave Square, London, 2 ed.

De Castro, E. and Morandi, C. (1987), 'Registration of translated and rotated image using finit Fourier transforms', *IEEE Trans. Pattern Anal. and Machine Intell.*, Vol. 9, No. 5, Pp. 700–703.

Dubois, E., Prasada, B. and Sabri, M.S. (1981), *Image Sequence Coding, in Image Sequence Analysis T. S. Huang Ed.*, Springer-Verlag, Berlin, Chap. 3, Pp. 229–287.

Duda, R.O. and Hart, P.E. (1972), 'Use of the Hough transform to detect lines and curves in pictures', *Communications of the Association for Computing Machinery*, Vol. 15, No. 1, Pp. 11–15.

Dudgeon, D.E. (1977), 'The computation of two-dimensional cepstra', *IEEE Trans. Acoust. Speech and Signal Processing*, Vol. ASSP-25, No. 6, Pp. 476–484.

Evens, R.G. (1983), 'Cost of digital angiography in 1983: Impact on planning and charges', In Felix, R. (Ed.), *Contrast Media in Digital Radiography*, Excerpta Medica, Netherlands, P. 396.

Feller, W. (1968), *An Introduction To Probability Theory and Its Applications*, Vol. 1, John Wiley & Sons, Inc., New York, 3 ed.

Feng, J. and Huang, T.S. (1984), 'Some experiments on estimating the 3-d motion parameters of a rigid body from two consecutive image frames', *IEEE Trans. Pattern Anal. and Machine Intell.*, Vol. PAMI-6, No. 5, Pp. 545–554.

Fischler, M.A. and Elschlager, R.A. (1973), 'The representation and matching of pictorial structures', *IEEE Trans. Computers*, Vol. C-22, Pp. 67–92.

Gabel, R.A. and Roberts, R.A. (1980), *Signal and Linear Systems*, John Wiley & Sons, Inc., 2nd ed.

Garden, K.L. (1984), *An Overview of Computed Tomography*, PhD thesis, University of Canterbury, New Zealand.

Gardenier, P.H., McCallum, B.C. and Bates, R.H.T. (1986), 'Fourier transform magnitudes are unique pattern recognition templates', *Biological Cybernetics*, Vol. 54, Pp. 385–391.

Garrett, J.A. and Smithson, P.H. (1987), 'Conventional x-ray imaging', *IEE Proceedings, Pt. A*, Vol. 134, No. 2, Pp. 107–114.

Geselowitz, D.B. (1979), 'Magnetocardiography: an overview', *IEEE Trans. Biomedical Engineering*, Vol. BME-26, No. 9, Pp. 497–504.

Gianino, P.D. and Horner, J.L. (1984), 'Additional properties of the phase-only correlation filter', *Optical Engineering*, Vol. 23, No. 6, Pp. 695–697.

Gibbons, J.D. (1971), *Nonparametric Statistical Inference*, McGraw Hill.

Gonzales, R.C. and Wintz, P. (1977), *Digital Image Processing*, Addition-Wesley, Reading, Massachusetts.

Goodman, J.W. (1968), *Fourier Optics*, M^{c}Graw-Hill, New York.

Goshtasby, A. (1985), 'Template matching in rotated images', *IEEE Trans. Pattern Anal. and Machine Intell.*, Vol. 7, No. 3, Pp. 338–344.

Gould, K.L., Mullani, N. and Wong, W. (1986), *Positron Emission Tomography, in* Cardiac Imaging and Image Processing, McGraw-Hill, New York, Chap. 13, Pp. 333–360.

Guthaner, D.F., Wexler, L. and Bradley, B. (1985), 'Digital subtraction angiography of coronary grafts: optimisation of technique', *Amer. Jour. Roentgenology*, Vol. 145, Pp. 1185–1190.

Harrington, D.P., Boxt, L.M. and Murray, P.D. (1982), 'Digital subtraction angiography: overview of technical principles', *Amer. Jour. Roentgenology*, Vol. 139, October, Pp. 781–786.

Harris, F.J. (1978), 'On the use of windows for harmonic analysis with the discrete Fourier transform', *Proc. IEEE*, Vol. 66, No. 1, Pp. 51–83.

Hartley, M.G. (1988), 'Chain–code to CCD Cameras – The Story of 25 years of road traffic studies at UMIST, Manchester', Seminar in Depart. of EE & E, Univ. of Canterbury, New Zealand, August.

Haykin, S. (1983), *Communication Systems*, John Wiley & Sons, New York, 2 ed.

Heel, M.V. (1986), 'Angular reconstruction: a posteriori assignment of projection directions for 3d reconstruction', *Ultramicroscopy*, Vol. 21, Pp. 111–124.

SEE ERRATA Heintzen, P.H. (Ed.) (1971), *Roentgen-, Cine- and Video-desitometry: Fundamentals and Applications for blood flow and Heart Volume Determination*, Thieme, Stuttgart.

Heintzen, P.H. (1986), *Digital Angiocardiography*, McGraw-Hill, New York, Chap. 11, Pp. 239–279.

Heintzen, P.H. and Bürsch, J.H. (Eds.) (1978), *Roentgen-, Video-Techniques for Dynamic Studies of Structure and Function of the Heart and Circulation*, Thieme, Stuttgart.

Heintzen, P.H., Brennecke, R., Bürsch, J.H., Lange, P., Malerczyk, V., Moldenhauer, K. and Onnasch, D. (1975), 'Automated video-angiocardiographic image analysis', *Computer*, Vol. 8, July, P. 55.

Heintzen, P.H., Brennecke, R. and Bursch, J.H. (1978), 'Computerized ECG-gated angiocardiography', In *Proc. of the VIII World Congress of Cardiology, 17–23 Sept.* Tokyo, *Ed. S. Hayase*, Tokyo Women's Medical College, Pp. 1124–1126.

Hill, D.R. (Ed.) (1975), *Principles of Diagnostic X-ray Apparatus*, Philips Technical Library.

Hohne, K.H. and Bohm, M. (1983), 'Processing and analysis of radiographic image sequences', In Huang, T.S. (Ed.), *Image Sequence Processing and Dynamic Scene Analysis*, Springer-Verlag, Berlin, Pp. 603–623.

Horn, B.K.P. (1977), 'Understanding image intensities', *Artificial Intelligence*, Vol. 8, No. 2, Pp. 201–231.

Horn, K.P. and Schunck, B.G. (1981), 'Determining Optical Flow', *Artificial Intelligence*, Vol. 17, August, Pp. 185–203.

Horner, J.L. and Gianino, P.D. (1984), 'Phase-only matched filtering', *Applied Optics*, Vol. 23, P. 812.

Hough, P.V.C. (1962), 'Method and means for recognizing complex patterns', U.S. Patent 3069654.

Huang, T.S. (Ed.) (1981), *Image Sequence Analysis*, Springer-Verlag, Berlin Heidelberg New York.

Huang, T.S. and Hsu, Y.P. (1981), 'Image sequence enhancement', In Huang, T.S. (Ed.), *Image Sequence Analysis*, Springer-Verlag, Chap. 4, Pp. 289–310.

Huang, T.S. and Tsai, R.Y. (1981), 'Image sequence analysis: motion estimation', In Huang, T.S. (Ed.), *Image Sequence Analysis*, Springer-Verlag, Chap. 1.

Huang, T.S., Bolstain, S.D. and Margerum, E.A. (1986), 'Least-squares estimation of motion parameters from 3-d point correspondences', *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* Miami Beach, FL, June 24–26.

Huber, P.J., Kleiner, B., Gasser, T. and Dumermuth, G. (1971), 'Statistical methods for investigating phase relations in stationary stochastic processes', *IEEE Trans. Audio Electroacoustics*, Vol. AU-19, No. 1, Pp. 78–86.

Hunter, G.J.S., Hunter, J.V. and Brown, N.J.G. (1986), 'Parametric imaging using digital subtraction angiography', *The British Journal of Radiology*, Vol. 59, January, Pp. 7–11.

Hurst, J.W. and Logue, R.B. (1978), *The Heart*, McGraw-Hill, New York, 4 ed.

Johns, E.J. and Cunningham, J.R. (1974), *The Physics of Radiology*, Charles C Thomas, Springfield, IL.

Justuson, B. (1981), 'Median filters: statistical properties', In Huang, T.S. (Ed.), *Two-Dimensional Digital Signal Processing II: Transforms and Median Filters*, Springer-Verlag.

Karbelkar, S.N. and Nityananda, R. (1987), 'Atmospheric noise on the bispectrum in optical speckle interferometry', *Jornal of Astrophysics and Astronomy.*, Vol. 8, Pp. 271–274.

Kashef, B.G. (1985), 'Image matching algorithms', In *Digital Image Processing*, *SPIE Proc. Vol. 528*, Pp. 14–22.

Kashef, B.G. and Sawchuk, A.A. (1983), 'A survey of new techniques for image registration and mapping', In *Applications of Digital Image Processing VI, SPIE Proc. Vol. 432*, Pp. 222–239.

SEE ERRATA Kim, Y.C. and Aggarwal, J.K. (1987), 'Positioning three dimentional objects using stereo images', *IEEE Journal of Robotics and Automation*, Vol. 3, August, Pp. 361–373.

Kruger, R.A. (1981), 'Estimation of the diameter of and iodine concentration within blood vessels using digital radiography devices', *Medical Physics*, Vol. 8, No. 5, Pp. 652–658.

Kruger, A.R. (1986), *X-ray Digital Cineangiography*, McGraw-Hill, New York, Chap. 2, Pp. 24–40.

Kruger, R.A. and Liu, P.Y. (1982), 'Digital angiography using a matched filter', *IEEE Trans. Medical Imaging*, Vol. MI-1, No. 1, Pp. 16–21.

Kruger, R.A., Mistretta, C.A., Crummy, A.B., Sackett, J., Goodsitt, M.M., Riederer, S.J., Houk, T.L., Shaw, C.G. and Fleming, D. (1977), 'Digital K-Edge subtraction radiography', *Radiology*, Vol. 125, October, Pp. 243–245.

Kruger, R.A., Mistretta, C.A., Houk, T.L., Riederer, S.J., Shaw, C.G., Goodsitt, M.M., Crummy, A.B., Zwiebel, W., Lancaster, J.C., Rowe, G.G., and Fleming, D. (1979), 'Computerized fluoroscopy in real time for noninvasive visualization of the cardiovascular system', *Radiology*, Vol. 130, Jaunary, Pp. 49–57.

Kruger, R.A., Mistretta, C.A. and Riederer, S.J. (1981), 'Physical and technical considerations of computerized fluoroscopy difference imaging', *IEEE Trans. Nucl. Sci.*, Vol. NS-28, No. 1, Pp. 202–212.

Kruger, R.A., Miller Jr., F.J., Nelson, J.A., Liu, P.Y. and Bateman, W. (1982), 'Digital subtraction angiography using a temporal bandpass filter: associated patient motion properties', *Radiology*, Vol. 145, November, Pp. 315–320.

SEE ERRATA Kruger, R.A., Bateman, W., Liu, P.Y. and Nelson, J.A. (1983), 'Blood flowdetermination using recursive processing: a digital radiographic method', *Radiology*, Vol. 149, Pp. 293–298.

Kuglin, C.D. and Hines, D.C. (1975), 'The phase correlation image alignment method', In *Proc. IEEE 1975 Int. Conf. Cybern. and Soc.*, Pp. 163–165.

Kuglin, C.D., Blumenthal, D. and Pearson, J.J. (1979), 'Map-matching techniques for terminal guidence using Fourier phase information', In *Digital Processing of Aerial Images*, SPIE, Vol. 186, P. 21.

Labeyrie, A. (1970), 'Attainment of diffraction-limited resolution in large telescopes by fourier analysing speckle patterns in star images', *Astronomy Astrophysics*, Vol. 6, No. 1, P. 85.

Labeyrie, A. (1976), 'High resolution techniques in optical astronomy', In Wolf, E. (Ed.), *Progress in Optics*, North-Holland, Amsterdam, The Netherland, Pp. 46–87.

Lantz, B.M.T., Foerster, J.M., Link, D.P. and holcroff, J.W. (1980), 'Determination of relative blood flow in single arteries: new video dilution technique', *Amer. Jour. Roentgenology*, Vol. 134, June, Pp. 1161–1168.

Lee, D.J., Krile, T.F. and Mitra, S. (1988), 'Power cepstrum and spectrum techniques applied to image registration', *Applied Optics*, Vol. 27, No. 6, Pp. 1099–1106.

Leo, T.K. and Gerson, G. (1979), 'Guidance system position update by multiple subarea correlation', *SPIE Proc. Vol. 186*, May.

Lesem, L.B., Hirsch, P.M. and Jordan, Jr., J.A. (1969), 'The kinoform: a new wave front reconstruction device', *IBM Journal of Research and Development*, Vol. 13, No. 2, March, P. 150.

Levin, D.C. (1984), 'Digital subtraction angiography: myths and reality', *Radiology*, Vol. 151, No. 3, P. 803.

Levin, D.C., Schapiro, R.M., Boxt, L.M., Dunham, L., Harrington, D.P. and Ergun, D.L. (1984), '(review) digital subtraction angiography: principles and pitfalls of image improvement techniques', *Amer. Jour. Roentgenology*, Vol. 143, September, Pp. 447–454.

Levine, M.D., Youssef, Y.M. and Ferrie, F. (1979), 'Cell movements: its characterization and analysis', In *Proc. IEEE Workshop on Computer Analysis of Time-Varying Imagery*, Philadelphia, U.S.A, April, Pp. 93–95.

Limb, J. and Murphy, J. (1975), 'Estimating the velosity of moving images in TV signals', *Computer Graphics, And Image Processing*, Vol. 4, No. 4, Pp. 311–327.

Liu and Huang, T.S. (1988), 'Estimation of rigid body motion using straight line correspondences', *Computer Vision, Graphics, And Image Processing*, Vol. 43, No. 1, Pp. 37–52.

Liu, C.Y.C. and Lohmann, A.W. (1973), 'High resolution image formation through the turbulent atmosphere', *Optics Communications.*, Vol. 8, No. 4, Pp. 372–377.

Liu, P., Kruger, R.A., Nelson, J.A., Miller, F.J., Osborn, A.G. and Wojtowycz, M. (1985), 'Digital angiography: matched filtration versus mask-mode subtraction', *Radiology*, Vol. 154, No. 1, Pp. 217–220.

Lohamnn, A.W. and Thum, C. (1983), 'Histograms of image sequence spectra', In Huang, T.S. (Ed.), *Image Sequence Processing and Dynamic Scene Analysis*, Springer-Verlag, Berlin, Pp. 350–363.

Lohmann, A.W. (1970), 'Comments about phase-only holograms', *Acoustical Holography*, Vol. 2, Chap. 14.

Lohmann, A.W. and Wirnitzer, B. (1984), 'Triple correlations', *Proc. IEEE*, Vol. 72, July, Pp. 889–901.

Macovski, A. (1983), *Medical Imaging Systems*, Prentice-Hall, Inc., Englewood Cliffs, N.J.

Maddison, F.E. (1977), 'Patient evaluation and other clinical considerations', In Rutherford, R.B. (Ed.), *Vascular Surgery*, Saunders, Philadelphia, Pp. 251–254.

Maher, K.P. and Thomson, K.P. (1988), 'Image noise analysis in digital fluorography', *Australasian Physical & Engineering Sciences in Medicine*, Vol. 11, No. 4, Pp. 143–146.

Martin, W.N. and Aggarwal, J.K. (1979), 'Computer analysis of dynamic scenes containing curvilinear figures', *Pattern Recognition*, Vol. 11, No. 3, Pp. 169–178.

McAinsh, T.F. (Ed.) (1986), *Physics in Medicine & Biology*—Encyclopedia, Vol. 1, Pergamon Press, Oxford, UK.

McCallum, B.C., Gardenier, P.H. and Bates, R.H.T. (1986), 'Invertible invariant transformations for robotic catalogues', In *Proc. of the Internat. Conf. on Future Computing Systems*, Depart. of EEE, Univ. Canterbury, Christchurch, New Zealand, Pp. 151–158.

Metherell, A.F. (1969), 'The relative importance of the phase and magnitude in acoustical holography', *Acoustical Holography*, Vol. 1, Chap. 14.

Mistretta, C.A., Crummy, A.B., Strother, C.M. and Sackett, J.F. (Eds.) (1982), *Digital Subtraction Arteriography: An Application of Computerized Fluoroscopy*, Year Book Medical Publishers, Chicago.

Mitra, S., Nutter, B.S., Krile, T.F. and Brown, R.H. (1988), 'Automated method for fundus image registration analysis', *Applied Optics*, Vol. 27, No. 6, Pp. 1107–1112.

Moodie, D.S., Yiannikas, J., Gill, C.C., Buonocore, E. and Pavlicek, W. (1982), 'Intravenous digital subtraction angiography in the evaluation of congenital abnormalities of the aorta and aortic arch', *Amer. Heart Journal*, Vol. 104, No. 3, Pp. 628–634.

Moores, B.M. (1984), *Physical aspects of digital fluorography*, The Hospital Physicists' Association, Chap. 2, Pp. 45–58.

Moores, B.M. (1987), 'Digital x-ray imaging', *IEE Proceedings, Pt. A*, Vol. 134, No. 2, Pp. 115–125.

Morse, P.M. and Feshback, H. (1953), *Methods of Theoretical Physics*, Vol. 1, McGRAW-HILL.

McDonnell, M.J. (1975), *Nonrecursive digital image restoration*, PhD thesis, University of Canterbury, New Zealand.

Murch, R.D., Quek, B.K. and McGreger, B.P. (1988), 'Determination of an object's orientation from silhouettes.', In *Proceedings of 25th National Electronics Conference of New Zealand).*

Nagel, H.H. (1981), 'Image sequence analysis: what can we learn from applications?', In Huang, T. (Ed.), *Image Sequence Analysis*, Springer-Verlag, Berlin, Pp. 19–213.

Nagel, H.H. (1983), 'Overview on image sequence analysis', In Huang, T. (Ed.), *Image Sequence Processing and Dynamic Scene Analysis*, Springer-Verlag, Berlin, Pp. 2–39.

Nagel, R.N. and Rosenfeld, A. (1972), 'Ordered search techniques in template matching', *Proc. IEEE*, Vol. 60, Pp. 242–244.

Nelson, J.A. (1985), 'Newer subtraction and filtration techniques', *Radiologic Clinics of North America*, Vol. 23, No. 2, Pp. 185–192.

Neumann, B. (1984), 'Optical flow', *Computer Graphics*, Vol. 18, No. 1, Pp. 17–19.

Nikias, C.L. and Raghuveer, M.R. (1987), 'Bispectrum estimation: a digital signal processing framework', *Proc. IEEE*, Vol. 75, July, Pp. 869–891.

Oppenheim, A.V. and Lim, J.S. (1981), 'The importance of phase in signals', *Proceedings IEEE*, Vol. 69, Pp. 529–541.

Oppenheim, A.V. and Schafer, R.W. (1975), *Digital Signal processing*, Prentice-Hall, New Jersey.

Oppenheim, A.V., Lim, J.S. and Curtis, S.R. (1983), 'Signal synthesis and reconstruction from partial Fourier domain information', *Journal of the Optical Society of America*, Vol. 73, No. 11, Pp. 1413–1420.

Parzen, E. (1960), *Modern Probability Theory and Its Applications*, John Wiley & Sons, New York.

Pearson, J.J., Hines, D.C., Goldsman, S. and Kuglin, C.D. (1977), 'Vidio rate image correlation processor', In *Proc. SPIE, Appl. of Digital Image Processing (IOCC 1977)*, Pp. 197–204.

Peters, T.M. (1973), *Image Reconstruction from Projections*, PhD thesis, Department of Electrical Engineering, University of Canterbury, Christchurch, New Zealand.

Powers, J., Landry, J. and Wade, G. (1970), 'Computer reconstructions from phase-only and amplitude-only holograms', *Acoustical Holography*, Vol. 2, Chap. 13.

Pratt, W.K. (1974), 'Correlation techniques of image registration', *IEEE Trans. Aerospace and electronics systems*, Vol. AES-10, No. 3, Pp. 353–358.

Ramachandran, G.N. and Srinivasan, R. (1970), *Fourier Methods in Crystallography*, Wiley-Interscience, New York.

Ratib, O. and Rutishauser, W. (1985), 'Recent developments of cardiac digital radiography', *International Journal of Cardiac Imaging*, Vol. 1, Pp. 29–40.

SEE ERRATA Riederer, S.J. and Kruger, R.A. (1983), 'Intraveous digital subtraction: a summary of recent development', *Radiology*, Vol. 147, June, Pp. 633–638.

Riederer, S.J., Brody, W.R., Enzmann, D.R., Hall, A.L. and Maier, J.K. (1983), 'Work in progress: the application of temperal filtering techniques to hybrid subtraction in digital subtraction angiography', *Radiology*, Vol. 147, No. 3, Pp. 859–862.

SEE ERRATA

SEE ERRATA Ro, D.W. (1985), 'Retrospective gating of masks in digital subtration imaging of coronary arteries', *Technical Report No. MIPG102, Department of Radiology*, University of Pennsylvania, USA.

Ro, D.W., Axel, L., Herman, G.T. and LeVeen, R.F. (1987), 'Computed masks in coronary subtraction imaging', *IEEE Trans. Medical Imaging*, Vol. MI-6, No. 4, December, Pp. 297–300.

Robbins, J.D. and Netravali, A.N. (1983), 'Recursive motion compensation: a review', In Huang, T.S. (Ed.), *Image Sequence Processing and Dynamic Scene Analysis*, Springer-Verlag, Berlin, Pp. 75–103.

SEE ERRATA Roberts, L.G. (1977), 'Machine perception of three dimentional solids', In Aggarwal, J.K., Duda, R.O. and Rosenfeld, A. (Eds.), *Computer Methods in Image Analysis*, IEEE Press, New York, Pp. 285–323.

Rocca, F. (1972), 'TV bandwidth compression utilizing frame-to-frame correlation and movement compensation', In Huang, T.S. and Tretiak, O.J. (Eds.), *Picture Bandwidth Compression*, Gordon and Breach, London.

Rosenfeld, A. (1984), 'Image analysis: problems, progress and prospects', *Pattern Recognition*, Vol. 17, No. 1, Pp. 3–12.

Rosenfeld, A. and Kak, A.C. (1982), *Digital Picture Processing*, Vol. 2, Academic Press, New York.

Rosenfeld, A. and VanderBrug, G.J. (1977), 'Coarse–fine template matching', *IEEE Trans. Systems, Man and Cybernet.*, Vol. SMC-7, Feb., Pp. 104–107.

Rossmann, K. (1969), 'Point spread function, line spread function and modulation transfer function. Tools for the study of imaging systems.', *Radiology*, Vol. 93, P. 257.

Round, W.H. (1982), *Quantitative Two-Dimensional Echocardiography*, PhD thesis, University of Canterbury, Christchurch, New Zealand.

Rutishauser, W., Simon, H., Stucky, J.P., Schad, N., NOseda, G. and Wellauer, J. (1967), 'Evaluation of roentgen cinedensitometry for flow measurement in models and in intact circulation', *Circulation*, Vol. XXXVI, December, Pp. 951–963.

Sampson, R.E. (1987), '3D range sensor-phase shift detection', *IEEE Computer*, August, Pp. 23–24.

Shaw, C.G. and Bassano, D.A. (1982), 'Characterization of measurement of noise properties for a digital radiography system', In *Application of Optical Instrumentation in Medicine X, Vol. 347*, SPIE, Pp. 24–33.

Shirai, Y. (1978), 'Recent advances in 3-D scene analysis', *Proc. Internat. Joint Conf. on Pattern Recognition 4th*, Pp. 68–94.

Sinton, A.M. (1986), *Contributions to astronomical and medical information processing*, PhD thesis, University of Canterbury, New Zealand.

Sklansky, J. (1978), 'On the Hough technique for curve detection', *IEEE Trans. Computers*, Vol. C-27, July, Pp. 923–926.

Skolnik, M.I. (1962), *Introduction to Radar System*, McGraw-Hill, New York.

Smith, H.C., Sturm, R.E. and Wood, E.H. (1973), 'Videodensitometric system for measurement of vessel blood flow, particularly in the coronary arteries, in man', *The Amer. Journal of Cardiology*, Vol. 32, August, Pp. 144–150.

Smith, F.W., Mostafavi, H., Steding, T.L. and Poulsen, R.S. (1979), 'Optimum windows for image registration', *SPIE Proc. Vol.186*, Pp. 56–60.

Suntheralingam, R. (1987), *Application of Digital Image Processing to Coronary Angiography*, Master's thesis, University of Canterbury, Christchurch, New Zealand.

Svedlow, M., McGillem, C.D. and Anuta, P.E. (1978), 'Image registration: similarity measure and preprocessing method comparisons', *IEEE Trans. Aerospace and electronics systems*, Vol. AES-14, Jan., Pp. 141–150.

Tian, Q. and Huhns, M.N. (1986), 'Algorithms for subpixel registration', *Computer Vision, Graphics, And Image Processing*, Vol. 35, No. 2, Pp. 220–233.

Tsai, R.Y. and Huang, T.S. (1981), 'Estimating three-dimensional motion parameters of a rigid planar patch', *IEEE Trans. Acoust. Speech and Signal Processing*, Vol. ASSP-29, No. 6, Pp. 1147–1152.

Tsai, R.Y. and Huang, T.S. (1984), 'Multiframe image restoration and registration', In Huang, T.S. (Ed.), *Advances in Computer Vision and Image Processing, Vol. 1*, JAI PRESS INC., Chap. 7.

Tsai, R.Y., Huang, T.S. and Zhu, W. (1982), 'Estimating three-dimensional motion parameters of a rigid planar patch, II: singular value decomposition', *IEEE Trans. Acoust. Speech and Signal Processing*, Vol. ASSP-30, No. 4, Pp. 525–534.

Ullman, S. (1979), *The Interpretation of Visual Motion*, MIT Press, Cambridge MA.

VanderBrug, G.J. and Rosenfeld, A. (1977), 'Two-stage template matching', *IEEE Trans. Computers*, Vol. C-26, No. 4, Pp. 384–393.

VanderLugt, A. (1964), 'Signal detection by complex spatial filtering', *IEEE Trans. Information Theory*, Vol. IT-10, P. 139.

Venot, A. and Leclerc, V. (1984), 'Automated correction of patient motion and grey values prior to subtraction in digitized angiography', *IEEE Trans. Medical Imaging*, Vol. MI-3, Pp. 179–186.

Venot, A., Lebruchec, J.F., Golmard, J.L. and Roucayrol, J.C. (1983), 'An automated method for the normalization of scintigraphic images', *Jour. Nucl. Med.*, Vol. 24, Pp. 529–531.

Venot, A., Lebruchec, J.F. and Roucayrol, J.C. (1984), 'A new class of similarity measures for robust image registration', *Computer Vision, Graphics, And Image Processing*, Vol. 28, No. 2, Pp. 176–184.

Vogel, R., LeFree, M., Bates, E., O'Neill, W., Foster, R., Kirlin, P., Smith, D. and Pitt, B. (1984), 'Application of digital techniques to selective coronary arteriography: Use of myocardial contrast appearance time to measure coronary flow reserve', *American Heart Journal*, January, Pp. 153–164.

von Spreckelsen, M. and Wolschendorf, K. (1985), 'A method to determine the instantaneous velocity of pulsatile blood flow from rapid serial angiographies', *IEEE Trans. Biomedical Engineering*, Vol. BME-32, No. 6, Pp. 380–385.

Waag, R.B. and Gramiak, R. (1976), 'Methods for ultrasonic imaging of the heart', *Ultrasound in Medicine and Biology*, Vol. 2, Pp. 163–170.

Weigelt, G.P. (1977), 'Modified speckle interferometry: speckle masking', *Optics Communications.*, Vol. 21, No. 1, Pp. 55–59.

Weigelt, G.P. and Baier, G. (1985), 'R136a in the 30 Doradus nebula resolved by holographic speckle interferometry', *Astronomy and Astrophysics Letters*, Vol. 150, No. 1, Pp. L18–L20.

Weigelt, G.P. and Wirnitzer, B. (1983), 'Image reconstruction by the speckle-masking method', *Optics Letters*, Vol. 8, No. 7, Pp. 389–391.

Wells, P.N.T. (1977), *Biomedical Ultrasonics*, Academic Press, New York.

Widrow, B. (1973), 'The "rubber-mask" technique', *Pattern Recognition*, Vol. 5, No. 3, Pp. 175–211.

Wirnitzer, B. (1985), 'Bispectral analysis at low light levels and astronomical speckle masking', *Journal of the Optical Society of America A*, Vol. 2, No. 1, Pp. 14–21.

Wolferts, K. (1974), 'Special problems in interactive image processing for traffic analysis', In *IJCPR-74*, Copenhagen, August, Pp. 1–2.

Wood, E.H., Sturm, R.E. and Sanders, J.J. (1964), 'Data processing in cardiovascular physiolory with particular reference to roentgen videodensitometry', *Mayo clin. Proc.*, Vol. 39, P. 849. SEE ERRATA

Wood, E.H., Ritman, E.L., Sturm, R.E., Johnson, S.A., Spivik, P., Gilbert, B.K. and Smith, H.C. (1972), 'The problem of determination of the roentgen density, dimensions and shape of homogeneous objects from biplane roentgenographic data with particular reference to angiocardiography', In *Proc. San Diego Biomed. Symp.* **11**, P. 3.

Wu, Q.X. and Bones, P.J. (1987), 'Removal of spike interference from digitized angiographic images', *Australasian Physical & Engineering Sciences in Medicine*, Vol. 10, No. 2, Pp. 96–101.

Wu, Q.X., Bones, P.J. and Bates, R.H.T. (1989), 'Translational motion compensation for coronary angiogram sequences', To Appear in September's Issue of *IEEE Trans. Medical Imaging.*

Yen, B.L. and Huang, T.S. (1983), 'Determining 3-D motion and structure of a rigid body using straight line correspondences', *In* Image Sequence Processing and Dynamic Scene Analysis *(T. S. Huang, Ed.)*, Pp. 365–394.

Yuen, C.K. and Fraser, D. (1979), *Digital Spectral Analysis*, CSIRO (Australia) and PITMAN Publishing Limited (U.S.A).

Ziedes des Plantes, B.G. (1932), 'A new method of differentiation in roentgenography (planigraphy)', *Acta Radiol.*, Vol. 13, Pp. 182–191.